REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
14.Aug.02

3. REPORT TYPE AND DATES COVERED
DISSERTATION

4. TITLE AND SUBTITLE
DECISIONS WITHIN COMPLEX SYSTEM: AN EXPERIMENTAL APPROACH USING THE STRATEGEM 2 COMPUTER GAME

5. FUNDING NUMBERS

6. AUTHOR(S)
LT COL BOIS J R

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
STATE UNIV OF NY ALBANY NY

8. PERFORMING ORGANIZATION REPORT NUMBER
CI02-116

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

20020829 020

14. SUBJECT TERMS

15. NUMBER OF PAGES
150

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

DECISIONS WITHIN COMPLEX SYSTEMS: AN EXPERIMENTAL APPROACH USING THE STRATEGEM-2 COMPUTER GAME

by

J. ROBERT BOIS

A Dissertation
Submitted to the University at Albany
State University of New York
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Information Science Ph.D. Program

Rockefeller College of Public Affairs and Policy

State University of New York

2002

For Patrick

. . . My little fighter

Table of Contents

List of Figures

List of Tables

Information Science Ph.D. Program

Rockefeller College of Public Affairs and Policy

The University at Albany, State University of New York

ABSTRACT

Decisions within Complex Systems: An Experimental Approach Using the STRATEGEM-2 Computer Game

by J. Robert Bois

In 1989, John Sterman published his seminal paper, *Misperceptions of Feedback in Dynamic Decision Making.* His misperception of feedback hypothesis deals with the difficulty people have in managing complex environments even when they purportedly have perfect knowledge and have perfect information about the system. Over the years, several authors have attempted to consider how the human failures, which are a prominent part of the misperception of feedback hypothesis, can be reduced. However, these authors have achieved mixed results in attempting to make improvements to human decision support. It is the purpose of the current research to provide meaningful decision support to managers of complex environments. Specifically, the research used the STRATEGEM-2 simulation game and purposely developed a decision support method designed to improve human performance within a complex system. The experiment required subjects to make a single decision within a dynamic system where the task involved feedback delays, nonlinearity of system processes, positive feedback loops, and multiple cues. The decision support included a decision rule and a newly developed game instruction designed to improve participant knowledge and information about the microeconomy of the STRATEGEM-2 simulation. Results of the research have discovered that the new instruction and the decision support rule produced significant results in improving decision making. Additionally, this research demonstrates that the lack of participant motivation levels can mask decision support interventions. Subjects with high self-assessed motivation outperformed those subjects with lesser motivation levels.

ACKNOWLEDGEMENTS

A project such as this cannot go without recognition of some very special interests and individuals that directly, and indirectly, contributed to the final outcome. First, a great deal of appreciation must go the United States Air Force for having the courage and foresight to enrich the education of one of its officers. Rest assured that the taxpayers money was well spent on this worthwhile sponsorship. Secondly, the support of a loving family was a requirement that was dearly met. Many late hours of study, countless weekends, and virtually no vacation time, were only some of the sacrifices made by the Bois family in seeing not only this project to finality, but for two years of coursework preceding the one year of dissertation work. Third, a great many thanks must be accorded to the dissertation committee: Professor David Andersen as chairman, and Professors John Rohrbaugh and Terry Maxwell. Each of these fine men where stalwarts in the planning, execution, writing, and termination phases of this project, and their assistance has proved invaluable. May the good Lord continue to watch over these fine men as they mentor other worthy candidates in their research. Finally, a special recognition goes out to Peter Otto, a dear friend who has commiserated with the author on innumerable occasions and has been a prime motivator to seeing this project, and degree, completed in record time. May our friendship continue to flourish and strengthen in the years to come.

Cordially,

J. Robert Bois

J. ROBERT BOIS, Lt Col, USAF

Chapter 1

INTRODUCTION

Over nearly two decades, John Sterman has researched and written about dynamic decision-making of participants using the STRATEGEM-2 computer simulation game. In a seminal work (1989a), along with several other accompanying articles (1985, 1987, 1989b, 1994), he established that a misperception of feedback in decision environments exists on behalf of participants because they fail to take into account delays between their own decisions and the dynamics of the simulation environment. Further, Sterman suggests that participants are operating with perfect knowledge of the system structure along with perfect information. However, given these "perfect" settings, participants consistently perform poorly. Sterman (1989a) developed a misperception of feedback hypothesis misperception of feedback hypothesis from his research. See Chapter 2, The STRATEGEM-2 Game, for an expanded description of the misperception of feedback hypothesis.

George Richardson and John Rohrbaugh (1990) essentially challenged Sterman's (1987, 1989a) findings by hypothesizing that if participants were given better cues to consider in the simulation environment they would perform better. They replicated the Sterman (1987, 1989a) study using a changed interface that incorporated revised cue designs. The results, unfortunately were not as predicted. They were mixed – one half of the participants improved their scores while the other half performed worse.

Edward Howie, Sharleen Sy, Louisa Ford, and Kim Vicente (2000), revisited Sterman's (1989a) misperception of feedback hypothesis and once again attempted to improve upon poor participant performance. Howie and others' (2000) approach was very similar to Richardson and Rohrbaugh (1990) with respect to how the simulation information was presented to the participant. Additionally, they expanded their focus to include measuring the level of environment knowledge possessed by each participant. This was an important step forward, in that the assumptions made by "perfect knowledge" had not been tested up to this point. Unfortunately, like the Richardson and Rohrbaugh (1990) experiment, Howie and others (2000) achieved mixed results. However, they did conclude that improving how information is presented to game players does result in improved game scores.

Problem Statement

It is possible that the above findings have not completely resolved the issues surrounding the misperceptions of feedback hypothesis. A major concern is that the misperception of feedback hypothesis puts undue emphasis upon the notion that participants have "perfect knowledge of the system structure along with perfect information." It is more than likely that this cannot be so, and it was demonstrated to some degree by Richardson and Rohrbaugh (1990) and Howie and others (2000). For example, Howie and others (2000) demonstrated that knowledge of the system structure was far from ideal before (and even after) the experiment had taken place. The Howie and others (2000) study, along with Richardson and Rohrbaugh (1990) also made valid criticisms on

how the participants of the Sterman (1987, 1989a) studies did not actually have perfect information at their disposal.

Therefore, the problem statement for the current research is: That the explanations of poor performance produced in the Sterman (1987, 1989a) studies may have been flawed, at least to the extent of the "perfect knowledge and perfect information" line of reason. It may be possible, then, to train or aide participants to be better performers.

Why is this Important?

The problem statement is important for the following reasons:

First, Richardson and Rohrbaugh (1990) point out that the information presented in the Sterman (1987, 1989a) studies is less than adequate. Specifically, they determined that in order for participants to make the most out of the information presented by the Sterman (1987, 1989a) simulation, they would require a certain degree of sophistication that most likely would not reside with the average player. In other words, information can be better presented, along with assistance for cue interpretation that should result in improved participant performance.

Second, Howie and others (2000) produce a convincing argument that Sterman (1989a) did not provide appropriate, or adequate, information on the computer display of his simulation. They suggest that substandard performance on behalf of participants is not due to a lack of knowledge or to psychological limitation.

Third, Howie and others (2000) demonstrated that the premise of "perfect knowledge" does not exist. Participants who were tested a priori and a posteriori exhibited knowledge that was far less than optimal.

Fourth, when preparing for this undertaking, Rohrbaugh[†] suggests that it has become evident that the "setup" of the experiment is equally crucial to the actual experiment itself. Apparently, and too often, researchers provide participants with instructions, and then, "jump" right into the data collection process. In other words, not enough attention has been paid to how participants are instructed. Is it possible then, that participants can be better prepared, or informed, regarding the dynamics of the simulation environment that they are about to take part? Possibly so.

Fifth, and most importantly, it is imperative to learn how to improve dynamic decision-making support. If indeed participants in an experimental setting can learn how to improve their performance with simulated complexity, then it may be possible to design decision support systems to assist real decision makers with the complexities they face in real systems.

Finally, one should not impart from this research that it is an attempt to debunk the misperception of feedback hypothesis. To the contrary, although the misperception of feedback hypothesis may be based, in part, on an incorrect assumption (that participants have perfect knowledge and information), it remains important to realize that human judges have difficulty with delayed feedback systems. Therefore, it is equally important to explore methods that can be used to improve human performance.

[†] 2000. Rohrbaugh, J. Personal interview. 4 January.

Purpose

In light of the above citations, it is the opinion of this researcher that it is warranted to try "once again" and see if participants can indeed perform better. There are too many "mixed results" requiring further/additional exploration.

Research Hypotheses

Assuming there are ways to improve human performance in the face of time-delayed feedback dynamics, the following hypotheses are projected for this research thesis. They are:

1. If information and knowledge about a system are better understood, participant performance will improve.
2. If participants are provided with a decision rule that focuses their attention on proper cues and how to weigh their importance, their performance will improve.
3. Participants reporting greater effort during the experiment simulation will out-perform those who do not.

Research Questions

Given the stated hypotheses, the following research questions are of close interest:

1. Can proper/adequate knowledge and information about the system be taught to participants?
2. Can participant performance be improved via decision cues and weights?
3. Can a participant's self-assessment of level of effort be used to better determine their own experiment performance?

Chapter 2

THE STRATEGEM-2 GAME

Background

In order to commence, the reader must first be made familiar with the mechanics of the STRATEGEM-2 game. The term STRATEGEM stands for a "**STRATE**gic **G**ame for **E**ducating **M**anagers." STRATEGEMs are a series of games produced for portable computers and were developed by the International Institute for Applied Systems Analysis (Laxenburg, Austria), the Resource Policy Center (Dartmouth College), and by the System Dynamics Group (MIT). STRATEGEM-2 deals with a micro-economy. It was born from the study of the economic long wave, or Kondratiev Cycle (Kondratiev, 1935). STRATEGEM-2 first appeared in the literature by Sterman and Meadows (1985) and was developed to teach decision-making dynamics to individual players (or teams) facing positive feedbacks inherent to a Kondratiev Cycle.

Game Play

Briefly, the game is played as follows: The player is established as a manager for a capital-producing sector of an economy. Game time is divided into two-year intervals beginning with year zero and ending in year seventy. Thirty-five decisions will be required from the player over the seventy-year period. The game board (see Figure 1 below), taken from the original Sterman experiments (1985, 1987, 1989a), is divided into two sectors, a capital sector (in simplicity, this would be the physical/industrial capacity to produce consumer goods and its own capital goods), and a goods sector (you may think of this as

the consumer sector). Orders for each sector go into a "backlog of unfilled orders" area where they will sit awaiting shipment to their respective sectors. The amount to be removed from this waiting area is equal to the capacity of the capital stock. Additionally, the capital stock loses ten percent of its level every two years due to depreciation.

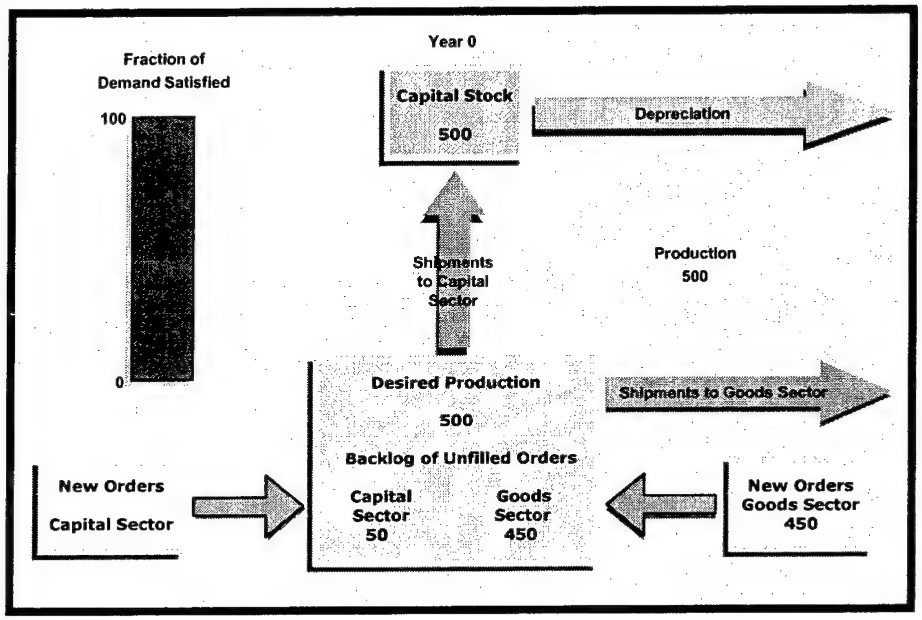


Figure 1 - Sterman STRATEGEM-2 Game Board

In the Sterman experiments (1985, 1987, 1989a), the game begins in equilibrium. This means that the capital stock is at a level of 500. The total for the backlog of unfilled orders is 500 as well (450 unfilled orders for the goods sector and 50 unfilled orders for the capital sector). Finally, a predetermined order of 450 goods sector orders is displayed for the player. This leaves a single decision to be made: How many orders are to be placed in the capital sector? The adept player should be able to recognize that an order of 50 for the capital sector will keep the game in equilibrium. The reason this is so is that 50 units to the capital sector will eventually be used to replace the 50 units of depreciation the capital

stock is scheduled to lose (500 * 10%). The combination of these 50 capital sector orders with the established 450 goods sector orders totals 500 units, which is equal to the production capacity of the capital stock. The capital stock will then be able to produce the 450 orders required by the goods sector, and it will be able to produce the 50 orders of capital to replace the 50 units it will lose to depreciation. Therefore, the game will remain in equilibrium.

On the upper left side of the game board is a "thermometer-type" display called, "Fraction of Demand Satisfied" (FDS) -- the bar indicates 100% FDS in year zero. Sterman (1985, 1987, 1989a) also produces a "Production" figure in order to provide information to the player. Production is calculated as the minimum of either capital stock or desired production. Plainly stated, industry would not produce more than demand requires. If the capital stock were larger than desired production, the player would simply be penalized for excess capacity. The FDS bar is merely a function of production divided by desired production. Hence, the only time FDS is less than 100% is when the capital stock is less than the desired production. Figure 2, below, depicts the STRATEGEM-2 microeconomy from a "stock and flow" perspective used by system dynamists to better show the feedback structure of the economy.

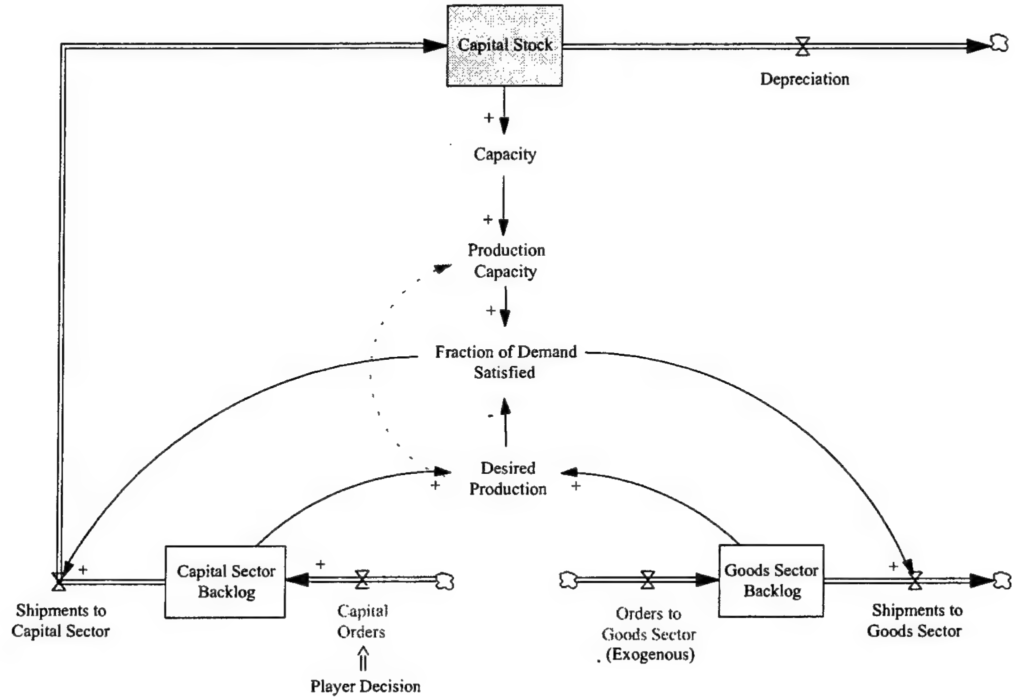


Figure 2 - STRATEGEM-2 Stock and Flow Structure

As a final note on the game mechanics, after each round of play, the game produces a "score" indicating to the player his or her level of performance. The score is a simple mathematical formulation that keeps an accumulating sum of the absolute difference between the total desired production (the total backlog of unfilled orders), and the production capacity of the capital stock (which is equal to the total of the capital stock), and is divided by the time interval (the years of play). For example, after the first round of play, the absolute difference between desired production and production capacity is zero. Divide that by 2 years and the score remains at zero. The score indicates how well each player can balance the interactions of supply and demand. There is equal penalty for excess demand, as well as excess supply.

Complexity Added

To provide complexity to the game, in year four, Sterman (1985, 1987, 1989a) adds a single step increase to orders from the goods sector. Orders go up from 450 to 500 and remain at that level for the remainder of the game (players in the game are unaware of this step increase, or of its longevity[†]). The key is that the participant must order more than the depreciation of the capital stock. The reason: The capital stock must be increased in order to meet capacity requirements for the new demand. The "gotcha" of the problem is that the player is more than likely unaware that it will take a few to several years to build up the capital stock to meet the new requirement. Additionally, the increased order to the goods sector further complicates the problem as it continues to grow the backlogs of orders that need to be shipped. This requires that more capital stock be ordered so that the demands of the burgeoning backlog of unfilled orders can be met – a classic multiplier/investment accelerator problem (see Figure 3 below). This positive reinforcing loop in the system normally forces players to order too much capital stock over subsequent years.

[†] This is a requirement of the experiment. Otherwise, if the subject were to know this information, he or she could possibly plan accordingly and defeat the dynamics the system is trying to simulate.

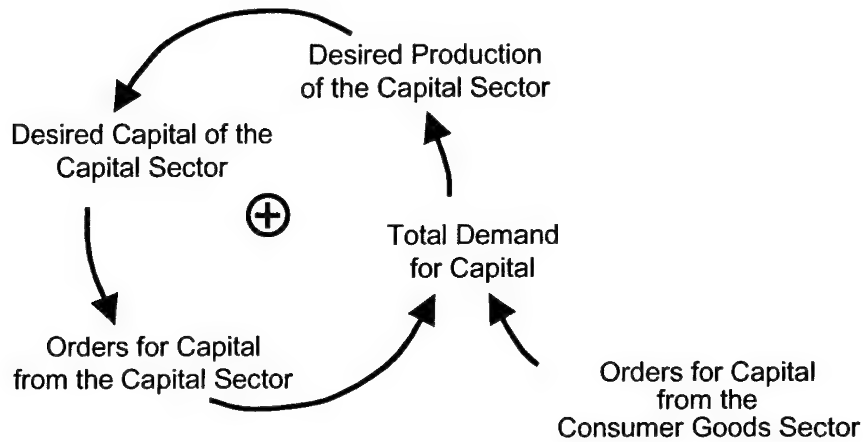


Figure 3 - Multiplier Accelerator Loop
Adapted from (Sterman, 1989a)

Typically, players fail to calculate that with the new increase in orders from the goods sector produces a new equilibrium (the actual new equilibrium raises from 500 to 555 – and will be presented as 560 in the game itself because the simulation rounds to the nearest 10). Therefore, as players build their capital stock to levels well above 560 (trying to counteract the increasing total backlog), they are slow to find that the backlog will quickly drop when capacity to produce is great and the orders for goods sector units remain at 500. When participants in the game realize that they now have too much capital sector inventory, they will tend to stop ordering all together. Depreciation then begins to show its effect by lowering the capital stock. However, the player soon finds him or herself "behind the power curve" once again with not enough capital stock to meet the total demand of the economy -- and the cycle continues. All of these problems are the result of poor anticipation of the delays in the system, as well as not calculating the desired new equilibrium level.

The Misperception of Feedback Hypothesis

The research performed by Sterman (1989a) has yielded a misperception of feedback hypothesis that attempts to capture why human subjects perform poorly in this simulation game. Simply stated, the misperception of feedback hypothesis occurs in two forms. The first is the *misperception of time delays*.

> "Failure to appreciate time delays is reflected in two distinct facets of the experimental results. First, there is a strong tendency for subjects to be overly aggressive in their attempts to correct discrepancies between the desired and actual capital stock [in the game]. Second, there is a strong tendency to ignore the time lag between the initiation of a control action and its full effect (Sterman, 1989a, pg. 324)."

The second form of the misperception of feedback hypothesis comes from *decisions to the environment*.

> Average behavior on behalf of participants, "would produce excellent results if demand were exogenous [to the system]. But demand is not exogenous. The multiplier feedback causes the environment to react endogenously to the decisions of the subjects. Their decision process, however, appears to be predicated on an exogenous environment. Thus many subjects were surprised that they did not receive all the capital they ordered as they tried to boost capacity. They were confused by the fact that placing orders to increase capacity seemed to worsen the gap between demand and supply. And they were further shocked that desired production suddenly dropped just when they thought they had finally caught up. These phenomena are direct consequences of the multiplier loop, that is, feedbacks from the subject's actions to the environment. In the long run, ordering more capital does increase capacity, but in the short run it adds to the total demand, worsening the shortfall. Ordering more capital also raises desired production further above capacity, reducing the fraction of demand satisfied and delaying delivery. During the period of

inadequate capacity, unfilled orders accumulate in the backlog, swelling desired production. When capacity finally overtakes desired production, these accumulated orders are shipped, and desired production falls (Sterman, 1989a, pg. 326)."

Therefore, the misperception of feedback hypothesis can be reduced to a subject's failure to appreciate the time delay built into the game, and that they fail to appreciate how their decisions are reflected within the game-playing environment.

Chapter 3

LITERATURE REVIEW

Overview

Dynamic Decision Making, or DDM, has been extensively written about in the literature for the past several decades. There have been several literature reviews written over this time covering the spectrum of DDM literature (Brehmer, 1992; Buchner, 1995; Funke, 1995; Hsiao, 1999; Kleinmuntz, 1987; Sterman, 1994). Of these, the Hsiao (1999) review of the DDM literature is probably the most comprehensive. He meticulously reviews and analyzes 33 DDM articles from the period of 1978 to 1998 (English language only).

Major Aspects of the DDM Literature

Hsiao (1999) examines the DDM literature from the similar perspective established by Funke (1995), which he categorizes relevant variables found in the DDM literature. Hsiao (1999) breaks the variables down into two major divisions. First, are the evaluative variables (read: dependent variables), and second, are the predictive variables (read: independent variables).

Dependent Variables

In the dependent variable division, Hsiao (1999) establishes five categories that the DDM literature has evaluated: Task performance, learning, efforts for decision making, quality of decision-making process, and decision-making architecture. Each of these

categories are further defined by sub-areas indicating the kind of measures that the literature was being represented in order to better define, or explain, each category.

Task Performance

In the first category, task performance, the overall idea is self-explanatory. Basically, how well do participants perform during the experiment. Hsiao (1999) uses five sub-areas of measures for this category: First, is the optimizing, maximizing, or minimizing, specified measures or benchmarks, second is reaching specified targets, then there are task systems behaviors, fourth are goals combining two criteria, and finally are goals combining greater than two criteria.

Of these sub-areas, it is the "optimizing, maximizing, or minimizing, specified measures or benchmarks" that are important to the current research. For example, studies that explore cost (the higher the cost, the lower the performance) are of particular interest (Diehl & Sterman, 1995; Howie, Sy, Ford, & Vicente, 2000; Richardson and Rohrbaugh, 1990; Sterman, 1987; Sterman, 1989a; Sterman, 1989b; Sterman & Meadows, 1985). These studies specifically consider the dynamic decision making inherent to the STRATEGEM-2 game/model. Subjects within these studies are required to minimize inventory cost (minimizing capital stock in relation to demand for the same capital).

Task-Related Knowledge

The second category, learning, relates to the task-related knowledge possessed by players in the gaming/simulation process. What is important is to determine the level of knowledge the participant has about the complexity involved in the decision-making

process before and after the experiment in order to determine if any learning about the complexity of the process has taken place.

Hsiao (1999) provides five measures for the learning category. First, there is performance on preferred tasks, second is number matching certain types of mental models. Next is the number of correctness of mental models aligned with heuristics and goals. Fourth, is to measure mean scores of pre-game and/or post-game questionnaires with regard to procedural task knowledge of the experiment. Finally, and of particular importance to the current study, is the measuring of mean scores of pre-game and/or post-game questionnaires with regard to declarative knowledge (Howie et al., 2000).

Effort for Decision Making

The third category for dependent variables is the effort for decision making. Although task performance and learning are certainly measurable in a relationship as an evaluative variable, an individual's effort toward the experimental task is another form of providing direct observation to the researcher. This category may be subdivided into three measures. First, is the amount of decision time (how long does it take to make a decision). Second, is the amount of information use for specific information items (is the participant using the information provided in the experiment). Third, is the amount of discussion among participants (do they seek each other's help when allowed by the experiment).

The inference made by these measures indicates that there is greater effort when there are longer decision times, or greater information use, or more collaboration among team members. What is missing from the literature, perhaps, is how effort can be measured from a self-assessment perspective. It is possible, therefore, to design a post-game survey that

can be used to determine how each individual self-assessed their own level of effort. Did the individual become bored with the repetitive tasks of the experiment? Did this cause him or her to rush to finish? Did they lose interest? Did this cause them to not pay close attention? These are the questions that should be asked in order to rule out that poor performance was indeed not influenced by the "lack of trying."

Quality of the Decision-making Process

The fourth category is related to the quality of the decision-making process. It has two measures that can be used to define the category. First, is the decision scope. This considers the number of different decision rules employed by the participant. And second, is the reliability of the decisions made. This involves the fluctuations of the decisions being made.

Decision-making Architecture

The final category for dependent variables is the decision-making architecture. This refers to the organization of how decision tasks have been embedded (Brehmer and Allard, 1991). The measure for such is represented by one thing, the delegation of decision making.

In summary, the five categories for dependent variables deal with performance, learning, effort, quality, and architecture. The current study of the STRATEGEM-2 game will specifically be concerned with the performance, learning, and effort aspects of the game. For a more detailed exposé of this variable, along with its associated categories and measures, please refer to Appendix A.

Independent Variables

Hsiao (1999) develops a research framework of a dynamic decision making model (see Figure 4). This logical framework can be used to trace the decision dynamics from the decision makers, through their game interfaces, into the realm of task systems and complexity and then back through the interface to the decision maker.

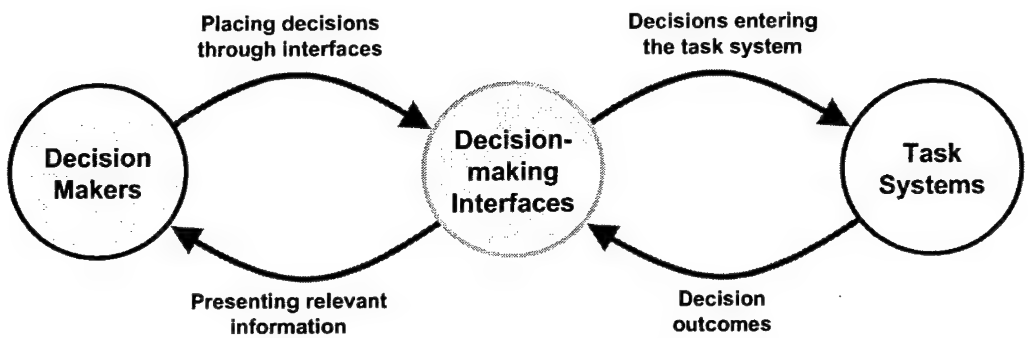


Figure 4 - Research Framework of Dynamic Decision Making
Adapted from (Hsiao, 1999)

Hsiao (1999) posits the formulation of the above model as follows:

> "Examining the predictors (independent variables) concerning the DDM research would be a logically subsequent task after pointing out those evaluative criteria (dependent variables). In doing so, Figure 4 may serve as a tentative framework containing four classes of objects and associated attributes. First of all, decision makers, with the attributes such as experience, cognitive style, intelligence, and expertise, should be an object of concern for they are the ones entering a series of decisions for a dynamic task. Then decision-making interfaces, in charge of entering the decision to the task system, provide decision makers with decision outcomes and relevant information that may help them to make the next decision. It is conceivable to

expect that the form and content of information display through the interfaces should matter in the DDM task. A task system contains task variables as well as their relationships and represents a problem with certain degree of complexity in the real world. A task system, usually programmed in a computer simulation game, produces decision outcomes. Surrounding the decision makers, decision-making interfaces, and task systems are decision-making environments, the setting in which decision makers may receive decision aids such as verbal instructions on task information and decision rules. Note that the decision aids are usually perceived through the decision-making interfaces. According to Figure 4, the current review categorizes various predictors of dynamic decision behavior extracted from the empirical studies into three groups: decision makers' factors, task complexity, and decision-making interfaces and environments (Hsiao, 1999, pg 9)."

Hsiao (1999) divides the independent, or predictor, variable into three broad categories. They are: Decision-maker factors, task complexity, and decision-making interfaces and environments. Each category is subdivided into conceptual definition forms with each of these forms being subdivided into various measures used by studies found in the literature (Appendix B).

Decision-maker Factors

The first category, decision-maker factors, refers to the intelligence, expertise, skills, and experience levels of decision makers. This category has four conceptual definitions. First, is cognitive style, operationalized by various personality tests such as Myers-Briggs Type Indicator, Gregoric Style Delineator, and Gordon's Cognitive Style Indictor. The

intent is to determine a relationship between differing personality types upon the evaluative variable. The second conceptual definition concerns task expertise and academic training. The idea suggests that experts can better handle complex situations than can novices. The third concept, computing skills, suggest that computing skill assists individuals to overcome the difficulties inherent to task systems. The final conceptual definition for the decision-maker factor category is practice and task experience. The body of evidence from the literature shows that performance and learning can indeed improve through familiarity and practice of certain tasks. Of some interest to this study is redundant testing (with complexity added) performed by Diehl and Sterman (1995). The measures used for the conceptual definitions described above can be found in Appendix B.

Another decision-maker factor that may be of importance, yet not found in the DDM literature, deals with cognitive dissonance theory. The theory, put forth by Festinger (1957), simply states that people are often comfortable with their cognitions of their surroundings, yet, when confronted with a cognition that is in direct conflict with one's own empirical cognition, a struggle arises within one's self that requires a resolution (General, 2002). The significance of this theory is that when one is put into the STRATEGEM-2 environment, they may in fact be dealing with competing cognitions that may possibly hamper performance.

Task Complexity

The second category for independent variables in the DDM literature is task complexity, that is, how do participants perform? The idea is that the more difficult the task, the poorer the performance will be. Therefore, what defines task complexity? There

are nine conceptual definitions for this category. They are explained as follows (refer to Appendix B for the outline of measures used):

1. Total number of variables: The more variables there are, the more difficult the task.

2. Interaction between subsystems: When variables have interaction effect among themselves (along with the dependent variable) increases complexity of the task.

3. Random variation: Variation among the variables increases difficulty.

4. Miscellaneous task characteristics: A Hsiao (1999) concept to capture other complexities not otherwise outlined here.

5. Time delay and lagged effects: This concept is most important to the STRATEGEM-2 simulation and is best captured by Hsiao (1999). "The DDM research has been focusing on additional complexity predictors mostly unique in dynamic task environments. *Lagged effects* of decisions refer to a common phenomenon that a previous decision at time t-1 does not always take effect immediately at time t. Subjects may not see the effect until time t+1 or later periods (e.g., Berry and Broadbent, 1988); or they may not even perceive the effect at all due to other task complexity factors. Lagged effects result from formulation, of which time delay is the most noticeable one (e.g., Sterman, 1989a, 1989b, Diehl et al., 1995) (pg. 10)."

6. Effectiveness of decisions on outcomes: These refer to the participant's ability to interpret lagged effects.

7. Frequency of oscillation: This refers to the stability of the system. The more unstable, the more difficult the task.

8. Positive feedback and gains: System dynamicists have been able to account for many of the instabilities observed in the previous concept through algebraic formulations of positive feedback loops within a system (e.g. Sterman, 1989a, Diehl and Sterman, 1995). In other words, decisions upon one or more variables will effect other parts of a system that will, in turn, affect future decisions. Another example is how "word of mouth" can effect positive gains in a market strategy game (Paich and Sterman, 1993).

9. Real-time simulation tasks: The body of work in this area deals with how complexity is added when the experiment is performed in a real-time, clock-driven, environment. Brehmer (1992), with his fire fighting experiments, are the cornerstone for real-time complexities in the DDM literature.

Decision-making Interfaces

The final category for independent variables is decision-making interfaces and environments. This captures the factors that are in between decision makers and task systems. Hsiao (1999) further defined this category into eleven conceptual definitions. Refer again to Appendix B for specific measures and studies applied against the following concepts:

1. Heuristics (decision rules) built into task systems: Rules that provide explicit instructions on what a decision should be based on previous outcomes. The Richardson and Rohrbaugh (1990) decision rule falls under this concept.
2. Modes of learning induced by lagged effects: "Specifically, when subjects perform a task without any lagged effects, they tend to concentrate on developing the relationships of the variables they think important. Comparatively, when subjects experience a task with lagged effects, they tend to be impressed with the cases of individually paired decision-outcome matches, without systematically forming variables' relationship. The former, termed the selective mode of learning (or explicit learning by other authors, e.g., Berry and Broadbent, 1988), enables subjects to acquire verbalized knowledge which can easily be measured by post-task questions. The latter, termed the unselective (or implicit) mode, may not be verbalized but still function in certain situations (Hsiao, 1999, pg. 12)." Additionally, this refers to suggesting that participants pay particular attention to key variables, instructions on variables' relationships, important feedback loops, and decision effectiveness that can be helpful (Berry and Broadbent, 1988; Wang, 1994).
3. Heuristics-induced goal setting that subjects receive through verbal instructions: These heuristics attempt to influence decisions, and therefore, performance and learning. The Sterman (1989a) decision rule is applied here.
4. Task property, strategies, and heuristics that subjects receive through verbal instructions: The importance here is to provide aids to participants to allow

them to understand task structure and the relationship among variables. The STRATEGEM-2 game studies performed by Richardson and Rohrbaugh (1990) fall into this precept.

5. Concurrent verbalization and thinking aloud: A concept that requires subjects to verbalize aloud their decision strategies and rules while making decisions. The idea is that through verbalization, participants will improve their task-related knowledge.

6. Increasing task salience: In order to increase task importance, Berry and Broadbent (1988) and Wang (1994) instruct participants about task structure information and how lagged effects may produce various outcomes. They found that their decision-aid supports task performance abilities.

7. Degree of decision precision required: In this concept, the participant is conditioned to learn better – to produce results at the "first decimal place versus the whole number."

8. Learning inducement: Produces subjects to search for better understanding of task structure over task performance.

9. Contents of information display: The exploration of this concept tests "what information is helpful?"

10. Forms of information display: Critical to the current study, this concept considers very closely the work established by Sengupta and Abdel-Hamid (1993), where they "base their research design on the theory of information

feedback and provide subjects with three types of computer information feedback: outcome feedback, cognitive feedback, and feedforward. Outcome feedback indicates online numerical reports for important state variables of the software project task. Subjects receiving cognitive feedback have access to online time plots containing the patterns of relevant variables and a tabular summary of these cues. Whereas outcome and cognitive feedback are always available on computer screens, feedforward is conveyed by an hour-long training session prior to the task, same as those decision heuristics described above (Hsiao, 1999, pg 14)."

"Decision rules and relevant cues have been incorporated in Richardson and Rohrbaugh's study (1990) where a group of subjects are provided with numerical weights of the three cues and the other group with the same plus a simple decision rule to transform the cues into a decision. Compared with the preceding instructions of decision rules and task property, the information display issue appears to be left unexplored (Hsiao, 1999, pg 14-15). This does not consider the initial foray into this area by Howie and others (2000).

11. Decision-making architectures: This concept purports that the command structure of a decision affects task performance. This can be formed as a networked architecture (open communication among participants), or a command-down architecture where participants receive orders from a single player.

For a more detailed exposé of the predictor variable, along with its associated categories and measures, please refer to Appendix B.

Original STRATEGEM-2 Studies

Sterman (1989a) examines human subject dynamic decision making using the STRATEGEM-2 computer game, and he bases this research on previous studies of his own from 1987 and 1985. Subjects are required to make a single decision on capital orders within a dynamic system. Tasks involve feedback delays, nonlinearity of system processes, positive feedback loops, and multiple cues. Sterman (1989a) viewed the problem as being framed by there being so little corporate epistemological[†] work relating decision-maker behaviors to large organization dynamics.

From the results of Sterman's (1987, 1989a) work, he develops a "decision rule" that purports to capture the decision-making behavior of people playing the game. From this, he simply recommends a three-cue task to his subjects in that they "order enough to replace depreciation, adjust it by some fraction of the discrepancy between the desired and actual levels of capacity, and don't forget to take the supply line of previous orders into account" (Sterman, 1987, p. 1588). He breaks this rule out into five different equations that logically capture how participants normally play the game.

Sterman (1989a) also displays an "optimal" solution to the game (Appendix C). A solution that has the game back into perfect equilibrium within five moves, or decisions (ten years of game time), following the step increase in goods orders in year four. A final

[†] From *epistemology*: A branch of philosophy that investigates the origin, nature, methods, and limits of knowledge.

score of 19 is produced using the "optimal" solution. Unfortunately, Sterman (1989a) does not provide the mechanics on how to determine the optimal solution.

Sterman (1989a) hypothesizes that his subjects would make decisions based on anchoring and adjustment heuristics as suggested by his decision rule and that they would be motivated by the observation that the complexity of determining the optimal rule (whatever that is) would be overwhelming to their abilities. In other words, Sterman (1989a) was predicting that players of the game would not perform well.

The results of his findings, from a field of 49 participants, produced a mean score of 591. However, the top half of the best participants had a score less than 300 (the top player produced a score of 77). Yet, as Sterman (1989a) observes, none were even close to the optimal score of 19. He was forthright when analyzing the (apparent) success of his experiment. He stated that: "The experimental results suggest that subjects do not behave optimally even when provided with perfect information and knowledge of the system..." Additionally, "the results reveal several misperceptions of feedback: many subjects fail to account adequately for the delay between their own decisions and the environment" (Sterman, 1989a, p. 329).

Early Critique of STRATEGEM-2 Findings

Richardson and Rohrbaugh (1990) posited the following: "How would players perform if the computer screen directly provided them with the cues appropriate for the task? What effect would different forms of cue presentation have on cognitive learning? These questions are important because they may reveal an alternative explanation for the misperceptions and dysfunctional behaviors found by Sterman (1989a). We hypothesize

that *the form* of cue presentation used for the study of decision making in dynamic environments will have a significant effect on results (pg. 464)." In order to analyze their hypothesis, they developed a three-condition experiment. Although the first two conditions of the experiment are important in their research design, they will not be discussed here in order that full attention can be given to the third condition, which represents the crux of their hypothesis.

The STRATEGEM-2 computer game was modified by Richardson and Rohrbaugh (1990) to accommodate and correct what they viewed as pitfalls of the Sterman (1987, 1989a) experiments. They felt that two cues, required for optimal play of the game, specifically, depreciation and shortfall, were not explicit on the (Sterman, 1987, 1989a) computer screen – they were assumed to be calculable by the player. Additionally, Richardson and Rohrbaugh (1990) suggested that players had to be sophisticated in order to use the Fraction of Demand Satisfied bar graph on the Sterman (1987, 1989a) screen. This is because players would have to realize that the "delay" or "production capacity" of their orders is the inverse of the FDS. They concluded that a better ordering strategy designed for the computer screen can be established.

The Richardson and Rohrbaugh Decision Rule

The game board used for the Richardson and Rohrbaugh (1990) experiment is replicated in Figure 5 (below). One can see that it has several changes from the one used in the Sterman (1989a) experiment (Figure 1). First, the depreciation of capital is explicitly shown (50 units). Second, production has been replaced with shortfall (the desired production minus the capital stock). Finally, a decision rule was put in place of the

Sterman FDS bar graph. The Richardson and Rohrbaugh decision rule specifies that the player would want to perform the following in year zero: 1) Take the current depreciation of 50 units and multiply it times 2 (for 100). Add to that, 2) Take the shortfall, currently 0 and divide by 2 (for 0), and finally, 3) Subtract the current capital backlog of 50. The rule proposed here by Richardson and Rohrbaugh (1990) would then prompt the player to order 50 units in year zero, which, of course, will keep the game in equilibrium.

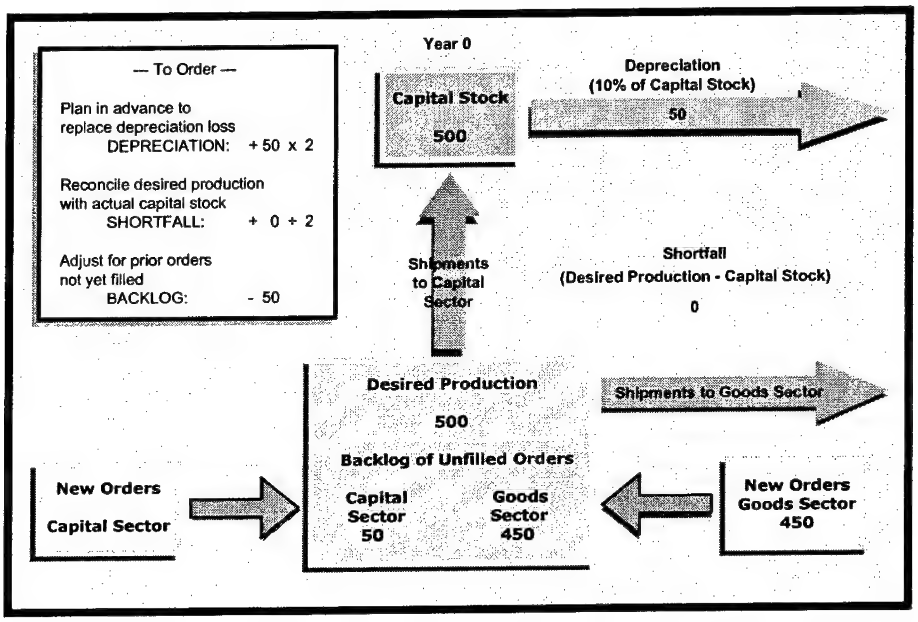


Figure 5 - The Richardson and Rohrbaugh SRATEGEM-2 Game Interface

Given the Richardson and Rohrbaugh decision rule, did their participants have to strictly adhere to this rule? No. As in the Sterman (1987, 1989a) experiment, players were always allowed to decide as best as they saw fit. The proposed rule provided by Richardson and Rohrbaugh (1990) was simply there to assist the decision makers in dealing with the complex task of balancing supply and demand in the micro-economy.

The Richardson and Rohrbaugh (1990) experiment had only 18 participants of which only 6 were randomly assigned to the condition described above. The authors received mixed results in their study. However, the results do not go without adequate explanation. For example, they were able to determine that one half of the participants were able to significantly perform to the level they had hypothesized. From the standpoint as an outside observer, it would additionally be concluded that the small sample size contributed to their sub-optimal findings. Richardson and Rohrbaugh (1990) did determine that when the subjects improved their consistency in information usage there is a link to maintaining stability in the system. Richardson and Rohrbaugh (1990), therefore, established that there is great room for additional research in this area.

Secondary Critique of STRATEGEM-2 Findings

In the Howie and others (2000) paper, the work of Richardson and Rohrbaugh (1990) was partially reexamined. The authors took to task the primary assumption of Sterman (1989a) that the players of the STRATEGEM-2 game have "perfect information" of the system. They viewed the misperception of feedback hypothesis included a certain degree of pessimism of the human endeavor within dynamic systems. From the most pessimistic angle, they quote Sterman as follows: "...improved performance can only be achieved through automated decision support because human dynamic decision making is bound to be poor because of 'a fundamental bound on human rationality'(Howie et al., 2000, pg. 152)." A less pessimistic examination of the misperception of feedback hypothesis is also presented in "that people do not do well because their knowledge of system structures is less than perfect. In this case, poor performance is caused by lack of knowledge rather than

some fundamental psychological limitation (Howie et al., 2000, pg. 152)." And finally, the authors conclude with a more optimistic approach by stating "that people perform poorly because the information they need to do the task is not presented in the computer display that they have available to them. In this case, the poor performance is caused by a lack of information rather than a lack of knowledge or a fundamental psychological limitation (pg. 152-153)."

The thrust of Howie and others' (2000) article is that players in the STRATEGEM-2 game do not have perfect knowledge or perfect information, as has been put forth in misperception of feedback hypothesis published studies. For example, according to Sterman (1987), participants had "perfect knowledge of the system structure and perfect information (pg. 1587)." To the contrary, Howie and others (2000), was the first to test player knowledge of the game itself. They provided a pre-game and post-game test covering the declarative knowledge of the simulation. Not one subject received a perfect score in either phase of testing. In fact, the test scores were rather low from the pre-game phase. Scores in the post-game phase did, however, show improvement over pre-game scores, which indicate that some knowledge had been gained from exposure to playing the game. The important item that Howie and others (2000) was showing was that participants did not have "perfect knowledge" of the game.

Similar to Richardson and Rohrbaugh (1990), Howie and others (2000) focused on the interface used during STRATEGEM-2 play. They developed a completely new Windows®-based interface. An interface that, in their argument, reflects proper design according to human-computer interaction principles that is in that body of literature. They

hypothesized that subjects playing the STRATEGEM-2 game with their newly designed interface, would perform better than participants playing the same game with the older interface developed by Sterman (1989a). See Appendix F for various displays of the Howie interface.

The results obtained by Howie and others (2000), like Richardson and Rohrbaugh (1990), were mixed. In the two trials that were performed, the group using the old interface performed better than the group using the newly designed interface, however, the standard error of their mean scores did overlap each other, which meant that some people in each group were actually performing on the same level as each other. In the second trial, the new interface group far outperformed the old interface group. This time, there was no overlapping of scores – the results were considered to be influenced by the new interface.

Other Pertinent Studies

There are two additional studies that have particular weight to the current research. First, is from Kim Vicente (1996), where he demonstrates that computer interface design can have positive, or negative, impacts on human performance. Second, the research by Sengupta and Abdel-Hamid (1993) purports that subjects provided with cognitive feedback, and subjects provided with feedforward show better performance than with subjects who are only provided outcome feedback information.

Vicente (1996) explored the possibilities of enhancing ecological interface design (EID) as a framework for complex human-machine systems. He proposed that the EID framework consisted of three principles (each intended to support a given level of cognitive control), they are:

1. "Knowledge-based behavior – this represents the externalized mental model that will support analytical problem solving
2. "Rule-based behavior – this provides a consistent one-to-one mapping between the work domain constraints and the cues provided by the interface
3. "Skill-based behavior – this supports interaction via time-space signals; the operator should be able to act directly on the display (pg. 253)."

In his study, Vicente (1996) simulated the dynamics of thermal-hydraulic process-control rooms (regulating the temperature and volumes of hydro-electric power facilities). The aim of decision makers is to keep water temperatures of two water reservoirs at a constant temperature, while at the same time, keeping each reservoir at a given water level.

Vicente (1996) acknowledged that in system dynamics studies, researchers have frequently concluded that participants are severely impaired in their ability to cope effectively with complex systems (for example: Sterman, 1989a). He posits that one potential explanation for this behavior is that research subjects did not have enough practice with a given simulation in order to adapt to its dynamics. Additionally, he further states another explanation in that, "the findings paint a very unflattering picture of people's capabilities to engage in dynamic decision making in complex systems which may, in part, be due to the impoverished interfaces used in those experiments. In fact, some authors explicitly refer to opaqueness (Brehmer, 1992) or lack of transparency (Dörner, 1987), as characteristics of dynamic decision-making problems. However, the research reviewed [in this study] shows that opaqueness is a property of an interface, *not* an inherent property of complex systems (Vicente, 1996, pg. 275)." And, "by accepting existing interfaces as an

unalterable given, one might be led to the unqualified conclusion that people are poor decision makers in complex systems (pg. 277)."

The work of Vicente (1996) was critical in developing a new STRATEGEM-2 interface as put forth by Howie and others (2000). This is also the reason the current research must incorporate the new interface design.

Kishore Sengupta and Tarek Abdel-Hamid (1993) develop alternative conceptions of feedback in dynamic decision environments: They experimentally investigate the relationships of outcome feedback, cognitive feedback, and feedforward in a system dynamics simulation game where human subjects perform duties as program managers by making decisions over the life of computer software projects. The concepts of each are as follows:

1. Outcome Feedback: "It has been argued that in dynamic environments, outcome feedback acquires the property of being corrective feedback in that it permits adjustments to the general direction of judgment (Hogarth, 1981). Decision makers can, therefore, rely on outcome feedback through a judgment-action-feedback loop to make effective decisions. Thus, a decision maker acting at time t has the benefit of outcome feedback from time $t-1$, enabling the individual to make appropriate changes in decision strategy (Sengupta and Abdel-Hamid, 1993, pg. 411)."
2. Cognitive Feedback: "Is conceptualized as information provided to a decision maker about (a) the relations in the decision environment, (b) relations perceived by the person about the environment, and (c) relations between the environment and the person's perception (Sengupta and Abdel-Hamid, 1993, pg. 412)."

Probably the most pertinent method for these conceptualizations to occur would be through task information and cognitive information. "Task information enables a decision maker to learn more about the environment, e.g. the relationship among the cues comprising a task. Cognitive information enables an individual to gain greater insight into his/her decision strategy, e.g. through information on weights accorded by the person to various cues (pg. 413)."

3. Feedforward: "Attempts to improve an individual's decision quality by providing him/her with a model of the task prior to performing the task[†]. . . Feedforward reduces the cognitive load on a subject because a large amount of the information the subject would have to infer from feedback is already transmitted through prior instructions. . . [The benefit of this method is that it] enables the decision maker to understand the key relationships and time lags that cannot be inferred from outcome feedback alone. Feedforward can thus serve as an effective method of planning an overall strategy (Sengupta and Abdel-Hamid, 1993, pg. 413-414)."

One of the largest problems facing people in dynamic environments deals with outcome feedback. Simply stated, outcome feedback does not provide enough information in order for decision makers to form adequate models of system behavior. The concepts of cognitive feedback and feedforward provide added insight and assistance to the decision maker. It is this researcher's position that cognitive feedback *along with* feedforward

[†] "The term feedforward is issued here in the specific sense of conveying task information to a decision maker, and should not be confused with the manner in which it is used by researchers in human-computer interaction (Sengupta and Abdel-Hamid, 1993, pg. 413)."

provides the best environment in which to make decisions when faced with complex systems. The Richardson and Rohrbaugh (1990) study inherently proposed such a design.

Designing a Tutorial Instruction Set

As suggested in Chapter 1, oftentimes researchers do not pay close enough attention to the experimental "setup" or instruction of human subjects. As will be seen in Chapter 4, Methodology, a computer-based on-screen tutorial is planned for the instructional phase of the experiment. In order to develop this portion of the experiment, a review of the literature is necessary in order to find common elements of good tutorial design.

Computer-based instruction, also known as: computer-aided/assisted instruction, multimedia interface, interactive multimedia instruction, and intelligent tutoring systems (just to name a few), is a relatively new phenomenon in educating persons and has experienced its largest developments in only the past two decades. Kemp and Dayton (1985) were early architects of computer-based instruction (CBI), and they recognized two distinct methods of teaching using a computer for tutorial purposes. The first of these methods presents information to the user in a fixed sequence. This "linear program" does not allow for differences among individual users. There may be some advantages to this type of tutorial programming. For example, if one wanted to research the individual differences in learning, this method would be more appropriate. The second method provides options to be chosen on behalf of the learner, allowing him, or her, to follow various paths of instruction. This is called a "branching program" and can result in better individualized learning.

Kemp and Dayton (1985) identified four major areas that CBI can be used to enhance individualized learning. They are:

1. Drill and practice: Provides practice for reinforcement of a concept or skill. The computer is used to provide a series of questions or exercises (similar to those found in a workbook).

2. Tutorials: They attempt to take the place of a human tutor. For example, they can be used to pose problems requiring a correct response from the participant, which navigates the learner to another block of instruction or to a remedial training block (depending on a correct or incorrect response).

3. Simulations: These are used to imitate dynamic processes or systems. For example, they can be used to navigate a sea-going vessel, manipulating an economy, or to operate the controls of some sort of manufacturing machine. In essence, simulations are used in order to provide the user an experience of real world without having to endure a real world consequence of a ship sinking, or an economic depression, or industrial accident. Additionally, these simulations significantly reduce training costs while at the same time, reduces the timeline required to gain the experience.

4. Games: If designed properly, games can be used to teach while taking advantage of the participant's competitive nature. The motivation to "win" in turn increases the learning outcome.

With regard to the current experiment, the intent is to develop the tutorial portion. Additionally, STRATEGEM-2 is not only a simulation, but it serves as a game as well. However, care must be taken when developing an instruction set for such an experiment. "In multimedia instruction, features of games and simulations are often combined, as both approaches offer highly motivational and potentially relevant environments. However, one caution must be underscored. Many simulations and games may not emphasize prescriptive instruction; the primary purpose of many games and simulations is entertainment or vicarious experience, with learning as a convenient by-product. Prescriptive instruction requires learning to be at the heart of the product, with the goals and parameters clearly defined (Schwier and Misanchuk, 1993)."

In developing a tutorial strategy, Halff (1988) espouses that the methods used to present material to the participant, depends on the instructional objectives and subject matter. He uses a dialogue strategy put forth by Collins and Stevens (1982) and is displayed in Table 1.

Instructional Objective	Strategies
Teach facts and concepts	Elicit fact or concept
Explain fact or concept	Teach rules and relations
	Case selection strategies
	Entrapment
Teach induction skills	Exercises and examples oriented to subskills

Table 1 - Tutorial Dialogue Strategies for Different Instructional Objectives

"Teaching of facts and concepts is accomplished by asking for or explaining the material. The decision to ask or tell is made on the basis of

the importance of the material and the student's knowledge thereof. Teaching of rules in tutorial sessions usually involves inducing the student to consider the relevant data and to formulate the rule. This can be done by presenting case data that makes the rule clear or by entrapment strategies that enable the student to eliminate incorrect versions of the rule. Skills for deriving [inducing] rules are taught as procedures. These procedures are broken down into their components (e.g., listing factors, generating cases to specification), and exercises and examples are provided that address each subskill (Halff, 1988, pg. 90)."

Soulier (1988) takes computer-based instruction to a higher level. He introduces the concept of management frames designed to aide the participant in his, or her, learning. Generically, CBI frames are considered important in the instructional process, however, they do not teach, per se. Rather, Soulier (1988) proposes "dialog frames" and "criterion frames" to better help the learner. "Dialog frames present information to the learner, as well as carry out an interactive dialogue/feedback between the learner and the computer. Criterion frames assess learning performance and provide feedback on results and follow-up activities (pg. 141)." Although not considered totally relevant to the current research problem, Soulier's insights may provide some direction towards development of an on-screen tutorial for participants of the STRATEGEM-2 game.

In developing computer-based instruction, attention must be paid to various display properties necessary to convey correct, and intended, information. One of the most critical to these properties is the passage length (Steinberg, 1990).

> "*Passage length.* The length of a passage is of special concern in CAI [computer-aided instruction]. When talking, the number of words used in an explanation is not restricted by space. In the classroom, an instructor may discuss a subject at some length, constrained only by time and his ability to maintain students' attention. Textbook writers may also present extensive discussions, limited only by publishers' page restrictions. In CAI, it is not feasible to present a great deal of verbal discourse. For some unexplained reason people read more slowly when text is presented on a display screen. Furthermore, students do not tolerate a computer program that is essentially an electronic page-turner. Perhaps this is because they are still unaccustomed to using a computer program for extended reading. It may be due to the expectation that a computer program should be highly interactive (Steinberg, 1990, pg. 84-85)."

The importance of Steinberg's (1990) "passage" prescription reflects upon the notion that the presentation must "get to the point." To belabor the participant with lengthy dialogs may detract, rather than, enhance the learning process. The important point for tutorial designer's to remember is try and capture specific learning points onto a single screen, or frame.

Although several authors have defined some of the necessary tools in developing various forms of CBI, Price (1991) brings the CBI designer in line with "goals and objectives" processes necessary to properly develop computer tutorials. For example, Price critically delineates that goals cannot be generalized; rather, they must be precisely stated

as clearly as possible (avoiding ambiguity that would leave questions in the mind of the learner/participant). Defining goals in terms of what the learner will actually do is also critical to this process (Price actually refines this prescription to stating goals in an active voice versus passive voice – "The learner will build a nuclear reactor," versus, "This lesson is about building nuclear reactors"). Objectives, simply stated, are used to indicate the performance of the learner.

The work of Roth and Hefley (1993) considers the technical perspectives of many investigations in intelligent multimedia presentation systems (IMMS). They review IMMS with regard to its purpose, key functional requirements, and architectural structure. They also consider the nature of information presented in various IMMS systems. Roth and Hefley (1993) describe two approaches to IMMS design. The first is a "task-analytic" approach that attempts to model actions, perceptions, and other cognitions on behalf of the IMMS user. The second is a "plan-based" communicative act view of an IMMS and emphasizes the presenter's goals.

There are several other applications within IMMS research that have similar approaches to the current study. For example, studies in factory management (Roth and Mattis, 1990, Gargan, Sullivan, and Tyler, 1988), financial models (Marks, 1991), marketing analysis (Anand and Kahn, 1992), project management (Roth and Hendrickson, 1991), and virtual worlds (Feiner, MacIntyre, and Seligmann, 1992).

Literature "On-ramp" to Study

In order to better focus on how previous studies of the STRATEGEM-2 literature applies to the current research, the following Figures 6 through 9 show how each of the

primary authors (Sterman, Richardson and Rohrbaugh, and Howie and others) have considered the various decision support factors that are currently in the in the dynamic decision-making literature. Additionally, the figures will show how the current author (Bois) has expanded upon the many various dependent and independent variables to be researched.

STRATEGEM-2 Literature Review
Dependent Variables

	Sterman	R & R	Howie	Bois
Performance	✓	✓	✓	✓
Optimization	✓	✓	✓	✓
Target Attainment		✓		✓
System Behaviors				
2-Goal Criteria				
> 2-Goal Criteria				
Knowledge			✓	✓
Declarative			✓	✓
Procedural			✓	✓
Correct Mental Models				
Matching Mental Models				
Transferred Tasks				

	Sterman	R & R	Howie	Bois
Effort				✓
Decision Time				✓
Information Use				
Discussion				
Self-Assessment				✓
Process Quality		✓		
Decision Scope				
Reliability		✓		
Architecture				
Delegation				

Figure 6 - STRATEGEM-2 Dependent Variable Summary
*Self-Assessment has been added by the author

STRATEGEM-2 Literature Review
Independent Variables

	Sterman	R & R	Howie	Bois
Decision-maker Factors		✓	✓	✓
Cognitive Style				
Expertise / Academic Training				✓
Computing Skill				
Practice / Task Experience		✓	✓	✓

Figure 7 - STRATEGEM-2 Independent Variable Summary (Decision-maker Factors)

STRATEGEM-2 Literature Review
Independent Variables

	Sterman	R & R	Howie	Bois
Task Complexity	✓	✓	✓	✓
Number of Variables				
Interaction Between Sub-systems				
Random Variation		✓		
Misc. Task Characteristics				
Time Delays	✓	✓	✓	✓
Decision Effectiveness				
Oscillation				
Positive Feedback / Gains	✓	✓	✓	✓
Real-time Simulation				

Figure 8 - STRATEGEM-2 Independent Variable Summary (Task Complexity)

STRATEGEM-2 Literature Review
Independent Variables

Interfaces / Environments	Sterman	R & R	Howie	Bois
Interfaces / Environments	✓	✓	✓	✓
Built-in Decision Rules / Heuristics		✓		✓
Learning via Lagged Effects				
Goal Setting Through Verbal Directions				
Decision Rules / Heuristics Verbally Given	✓			
Concurrent Verbalization				
Increasing Task Salience				
Precision Requirements				
Learning Inducement				✓
Information Display Content	✓	✓	✓	✓
Forms of Information Display		✓	✓	✓
Architecture				

Figure 9 - STRATEGEM-2 Independent Variable Summary (Interfaces / Environments)

Literature Summary

From the literature, it is recognized that STRATEGEM-2 is a very difficult game to play because of time delays, nonlinearities, and positive feedback loops. In determining the reason (or reasons) why people do poorly in the exercise has been the subject of much disagreement, particularly when it comes to suggesting methods for making improvements.

What is evident from the literature is that first, the forms of cues are important to the successful decision making faculties of participants (Richardson and Rohrbaugh, 1990; Vicente, 1996; Howie et al., 2000). Second, information presented to participants can be enhanced through the interface being used (Richardson and Rohrbaugh, 1990; Vicente,

1996; Howie et al., 2000). Third, participant knowledge is related to participant performance (Howie et al., 2000). Fourth, using an interactive on-screen tutorial may improve an individual's knowledge and perception of the micro-economy that in turn may improve game performance. And, finally, a self-assessment of "effort" may have bearing on the results observed.

Based upon the discoveries of the literature review, this study turns toward reaccomplishing major portions of studies already undertaken. Specifically, by those of Richardson and Rohrbaugh (1990), and Howie and others (2000).

Chapter 4

METHOD OF STUDY

Overview

The Richardson and Rohrbaugh (1990) study, along with Howie and others (2000), had very sound theoretical foundations in challenging a portion of the misperception of feedback hypothesis. Both studies attempt to fill gaps that are perceived to be unexplained in the misperception of feedback hypothesis. Their hypotheses, whether implicit (Richardson and Rohrbaugh, 1990), or explicit (Howie et al., (2000), stated that players in the STRATEGEM-2 game do not have perfect knowledge of their environment, nor does the environment display perfect information. Sterman (1989a) would argue that because the participant can view a graphic screen at anytime during the experiment, they could obtain immediate outcome feedback of what has been occurring in the dynamics of the game. Richardson and Rohrbaugh (1990) provided the exact same outcome feedback as well as provided current depreciation and shortfall information on the game board. Howie and others (2000), out-doing their predecessors, provided all this information on a single game-screen.

However, Howie and others (2000) explicitly brought up a very important facet of the misperception of feedback hypothesis worthy of further consideration – the finding that players, before and after the game, could not demonstrate "perfect knowledge" of the game. Richardson and Rohrbaugh (1990) also grappled with this aspect of Sterman's (1989a) work by asking: "How would players perform if the computer screen directly

provided them with the cues appropriate for the task? What effect would different forms of cue presentation have on cognitive learning? These questions are important because they may reveal an alternative explanation for the misperceptions and dysfunctional behaviors found by Sterman (1989a) (Richardson & Rohrbaugh, 1990, pg. 464)."

The fact that Richardson and Rohrbaugh (1990), and Howie and others (2000), arrived at mixed results (however, encouraging), leaves the issues of "perfect knowledge," modern computer interfaces, feedforward cues, and the cognitive learning processes to be unresolved. Therefore, the current study retested the Richardson and Rohrbaugh (1990) precepts. Additionally, the Howie and others (2000), computer interface was used along with the concept of testing participant knowledge.

Another issue that has been brought to the attention of the researcher is by Rohrbaugh. It refers to the "setup" of the experiment to the participants. Too often, according to Rohrbaugh, researchers preoccupy themselves with measuring the many dynamics of their experiments. They bring in human subjects, give them something to read, and then move into the experiment without ever considering whether the setup may have had an impact on the subject's performance. The concept of the setup in the STATEGEM-2 game can possibly have significant feedforward effect. Hsiao (1999) found a study where this procedure was introduced as a distinct form of measurement. "Sengupta and Abdel-Hamid (1993) base their research design on the theory of information feedback and provide subjects with three types of computer information feedback: outcome feedback, cognitive feedback, and feedforward. Outcome feedback indicates online numerical reports for important state variables of the software project task. Subjects receiving cognitive

feedback have access to online time plots containing the patterns of relevant variables and a tabular summary of these cues. Whereas outcome and cognitive feedback are always available on computer screens, feedforward is conveyed by an hour-long training session prior to the task (Hsiao, 1999, pg. 27)."

Reading the instructions to STRATEGEM-2, used by the three main studies identified in this paper, many questions remained unanswered. Therefore, improvements were attempted to the setup of the experiment that can transfer the dynamics of the game into meaningful knowledge that participants could better grasp and understand.

Furthermore, the sample size has been increased up to seven-fold from what the previous research populations have been. For example, the Richardson and Rohrbaugh (1990) study had 18 participants divided into 3 different conditions of 6 people each. Howie and others (2000) had 20 participants divided evenly into 2 treatment groups. The fact that neither study was able to produce meaningful results may be attributable to small sample sizes. The current study had a useful survey sample of 138 participants.

Design Proposal and Matrix

The following research proposal and matrix was used:

1. Used the old Sterman instructions in the control conditions presented by Howie and others (2000), (Appendix D)
2. Developed an on-screen tutorial to train game participants (Appendix E)
3. Used the Howie STRATEGEM-2 interface (Appendix F)
4. Tested game knowledge among the participants following train-up (Appendix G)

5. Surveyed the participants to determine their level of effort at the end of the experiment (Appendix H)
6. Performed a practice trial, and then two scored trials where orders to the goods sector remained the same (a single step increase in year four) for each trial
7. Enrolled 150 volunteer participants

8. Randomized participants into 4 treatments and conditions as follows:

    a. Receives an on-screen train-up of the Sterman instructions (presented by Howie et al., 2000), a practice trial, a knowledge survey, Q&A, and two measured trials.

    b. Receives an on-screen train-up of the Sterman instructions (presented by Howie et al., 2000), a practice trial, a knowledge survey, Q&A, the Richardson and Rohrbaugh decision rule, and two measured trials

    c. Receives new on-screen tutorial (Bois instructions), a practice trial, a knowledge survey, Q&A, and two measured trials

    d. Receives new on-screen tutorial (Bois instructions), a practice trial, a knowledge survey, Q&A, a practice trial, the Richardson and Rohrbaugh decision rule, and two measured trials

The above treatments and conditions are further explained by the following 2 x 2 matrix shown in Figure 10 (below). Along the vertical axis, there are two treatments that received (hypothesized) inadequate/adequate training (the original Sterman instructions, or better, not receiving the Bois Instructions, and receiving the Bois Instructions). The horizontal axis has two treatments that received (hypothesized) non-decision/decision support (no Richardson and Rohrbaugh rule and receiving the Richardson and Rohrbaugh Rule). Four conditions are created from the combinations created by mixing different levels of decision support and game instructions.

Conditions and Treatments	No Rule	Receives R & R Rule
No Bois Instruction	I	II
Receives Bois Instruction	III	IV

Figure 10 - Human Subject Random Group Assignments

Data

The "final score" produced by the simulation game was the main data point captured in this research. The score represented the subject's ability to manipulate capital sector orders in order to minimize overall backlog orders against the simulation's presentation of supply and demand, as well as minimizing overproduction of capital and goods sector orders. The score was determined by the average absolute deviation between the supply and demand for capital over the length of the game. The ultimate goal for the participant was to minimize his or her score – the smaller the score, the better the performance.

Additional data points collected included: The scores from the participant knowledge surveys as tested by Howie and others (2000), (Appendix G), and the determination of

participant level of effort from a self-assessment perspective (Appendix H). Additionally, an abundance of demographic information was collected and analyzed.

Sample and Subjects

Although the planned research will make inferences from the sample to the greater population, the researcher used a non-probability/convenience sample of human subjects. Specifically, participants were drawn from graduate/undergraduate students enrolled in the public administration, information science, business administration, finance, and marketing programs at the State University of New York at Albany. Student participants were offered a substitute option for other required course requirements in order to generate interest in the experiment. In total, 54 graduate and 96 undergraduate students elected to participate in this research.

Clearly, this sample is convenient to the researcher, yet it is also purposive in nature – these students where chosen because of who they are, and for the positions for which they are professionally preparing themselves. These students are getting ready for careers that will involve decision making that will be rooted in complex systems.

All subjects were recruited on a voluntary basis and did so without receiving any stipend. The participants were randomly placed into the treatments and conditions shown in Figure 10 above. Because the actual experiment was designed to last about two hours, several experiment periods were planned to accommodate the many and various schedules of the proposed participants. There were morning, mid-day, afternoon, and evening sessions.

Variables - Measures

The dependent variables measured in this experiment include the following: Score received on Trial 1, Score received on Trial 2, Mean average score for both trials, the change in scores between the first and second trials (obtained by subtracting Trial 2 from Trial 1), and the self-assessed level of effort. Independent variables included the game instruction setup, decision support, game knowledge, and demographic information.

Dependent Variable

During the experiment, the dependent variables were the scores received in the first and second trials, the mean average of both trials, the change in scores between the two trials (as a reminder, the lower the score, the better the performance for a given trial), and the self-assessed level of effort. The scores indicate the participant's ability to ferret out the important factors in the decision-making process within the dynamic system. Operationalization of this variable was derived through the actual decision process required by the participant to manipulate the computer simulation. For each of the trials in the experiment, the participant was faced with a total of 35 decision frames that spanned a total of 70 years. As the participant worked through the several decision frames, his or her individual decision scores were accumulated into a final score.

The participant level of effort was surveyed via a self-assessment (Appendix H). This dependent variable has three subsets: self-assessment of individual interest in the research, task understanding, and performance. These subset variables were operationalized through the various statements found in the survey instrument (Appendix H).

Following is the assignment of statements to each of the three subset variables:

Variable	Survey Statement Number
1. Self-assessment of performance:	3, 7, 8, 12
2. Self-assessment of research interest:	4, 6, 10, 11
3. Self-assessment of task understanding:	1, 2, 5, 9

The questions are designed to get the participants to accurately document their perceptions about their own actions during the experiment. It was presumed that the variables, when analyzed, would reflect on whether they had any significant predictability upon the dependent variables. The survey instrument is based upon a Likert-type scale. It is used to measure the internal states of the subjects (such as attitudes, emotions, and orientations) (Bernard, 2000). The design of each question in the instrument was used to measure whether the participants fully employed themselves during the experiment.

Independent Variables

Independent variables included the game instruction setup, decision support, game knowledge, and demographic information. Operationalization of these variables was as follows: Game instruction setup was derived from subjects receiving either the original Sterman instructions (Appendix D), or the newly devised Bois instructions (Appendix E).

Decision support occurred in two forms. In the first, participants received, or did not receive, the Richardson and Rohrbaugh decision rule. In order to produce this part of the experiment, a card with specific information about the decision rule was provided to those participants destined to receive decision support (see Figure 11 below).

----Front Side----

As the manager of the STRATEGEM-2 economy, you have taken it upon yourself to hire a very reputable economic consultant to assist you with your decisions. This person has determined that if you are to follow the formula in the blue box on the reverse side of this card, you will most likely receive an outstanding score for the game. You are reminded by this professional that although you are not required to heed the advice given, you <u>must</u> remain patient and diligent with using the formula.

Example on using the decision aide **<u>in year zero</u>** of the game:

1. Take the current depreciation of 50 units and multiply it times 2 (for 100).
2. Add to that the shortfall* (currently 0) and divide by 2 (which equals 0).
3. Then subtract the current <u>capital backlog</u> (*not total backlog*) of 50.
4. This produces an order of 50 capital units for Year 0

* Shortfall = (total backlogs - current capacity).
If this figure computes to less than zero, use zero.

----Back Side----

--- To Order ---

1. Plan in advance to replace depreciation loss

(DEPRECIATION x 2) = ______

2. Shortfall: Reconcile total backlogs with current capacity

add (SHORTFALL ÷ 2) = +______

3. Adjust for prior orders not yet filled

subtract (CAPITAL BACKLOG) = - ______

Total Orders = ______

Shortfall = (total backlogs - current capacity).
If this figure computes to less than zero, use zero.

Figure 11 - Richardson and Rohrbaugh Decision Rule Input Card

In the second form of decision support, participants received, or did not receive, the Bois instruction (Appendix E). This instruction was designed as an on-screen tutorial and had two learning inducement objectives in mind during development, 1) to get participants to understand how STRATEGEM-2 is played, the different features of the game board, and what information is being conveyed to the participant from the various features of the interface, and 2) to get subjects to understand the concept of "equilibrium" within the game dynamics. The tutorial was set-up as a linear program to introduce the various teaching elements and included a branching design as each successive page of the tutorial unfolded for the participant. Additionally, criterion frames were used to examine/test participant knowledge of the equilibrium concept and provided direct feedback in order to assist in the learning process. As a final note, close attention was paid to the passage lengths of each of the tutorial's pages so not to overtax subject attention spans.

Game knowledge was tested by adapting the Howie and others (2000) knowledge survey (Appendix G). Scoring of this survey was based upon the number of correct responses on a 0 to 100 percentage scale.

Regarding demographics, operationalization of this variable included participant: gender, age, graduate status, years of professional experience, total time on task, and test scores (knowledge survey).

Experiment Setup

For all conditions surveyed, the setup of the experiment included either the original Sterman instructions (Appendix D) or the Bois instructions (Appendix E) along with an overview of the Howie STRATEGEM-2 interface (see Figure 12 below). Additionally, the

conditions either included, or did not include, the Richardson and Rohrbaugh decision rule. After a train-up session was conducted, a practice trial of the game was played by each subject followed by the knowledge survey and a question and answer session.

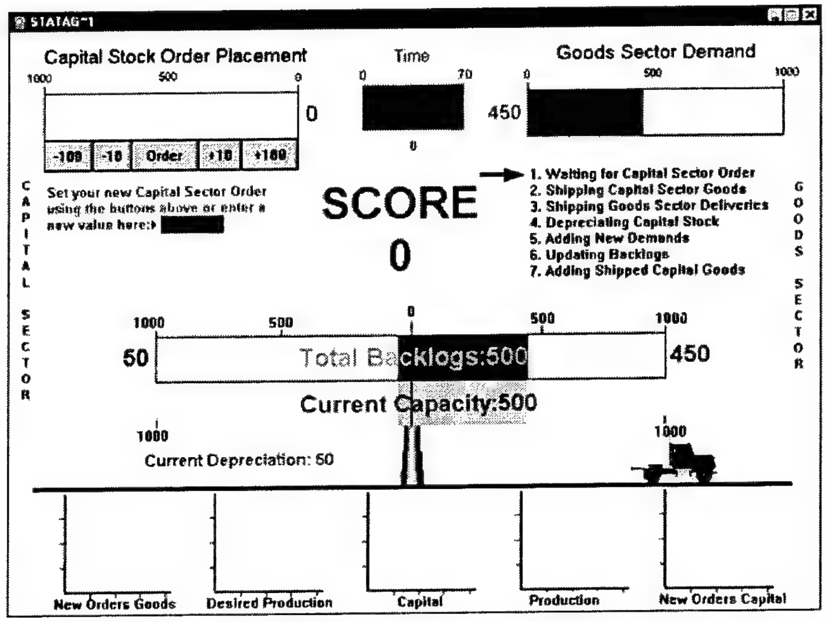


Figure 12 - Howie STRATEGEM-2 Interface

Data Collection Procedure

In all, 150 graduate and undergraduate students at the University at Albany volunteered to participate in the experiment. Each participant was randomly assigned to the various treatments of the experiment.

The first step of data collection process was to collect knowledge survey (test score) data. It measured the depth of participant knowledge of the STRATEGEM-2 system. This survey, or test (Appendix G), was administered following the train-ups of all participants. The survey was scored by the number of correct answers based on a 0 to 100 percentage scale.

After the setups for all participants were completed, along with their knowledge surveys, each group underwent two scored trials of STRATEGEM-2 using the Howie interface. This phase of the experiment collected individual participant data for the Trial 1 and Trial 2 final scores, the mean average score for the two trials, and the change in scores by subtracting Trial 2 from the Trial 1 score.

Following game play, all participants were administered the post-experiment written survey as follows: Upon completion of the gaming simulation, participants were presented with the self-assessment survey instrument and briefed about its contents, as well as about researcher expectations.

Data Analysis

All statistical analyses for this research were performed using the Statistical Package for the Social Sciences software (SPSS). The data analyses included simple descriptive statistics that were used to capture the broad spectrum of data points among the participants.

The main analysis performed was a 3-way analysis of variance. It was used for comparison of the instruction set (receives Sterman instructions or receives the Bois instructions) put against the decision support rule (receives or does not receive the Richardson and Rohrbaugh decision rule), and further compared with a measure of self-reported motivation. The analysis was used to determine the main effect of either the Bois instruction, the Richardson and Rohrbaugh decision rule, and motivation level upon participant performance.

Data Reduction

When considering the entire data set after all collection had been completed, it became obvious that some scores obtained in the two recorded trials were so high, that some form of reduction, or elimination of cases, would be required in order to better capture the true performance of the body of participants, and attempt to reduce or eliminate problems associated with regression to the mean. For example, over two thirds of all participants scored less than 1,000 points for either Trial 1 or Trial 2 (Reminder: The lower the score the better the performance. The Sterman optimal score for the game is 19, and the Richardson and Rohrbaugh decision rule produces and optimal score of 67). Additionally, three subjects scored in excess of 10,000 points in both Trials 1 and 2.

The researcher has determined that individuals receiving very high scores possibly did not understand the task, or they failed to grasp the requirements of the instructions. In order to set some sort of demarcation, any case with a Trial 1 or Trial 2 outlier in excess of 4,000 points was eliminated from the data set. Therefore, 12 cases were eliminated from the original 150, reducing the total N to 138, or by 8 percent.

Data Conversion

In order to better visualize the score data obtained during the experiment, Figure 13 shows boxplots for Trial 1 (T1), Trial 2 (T2), and the two-trial average (TA) for the four conditions generated from the various treatments. Additionally, it demonstrates the existence of several mild and extreme outliers of the raw scores.

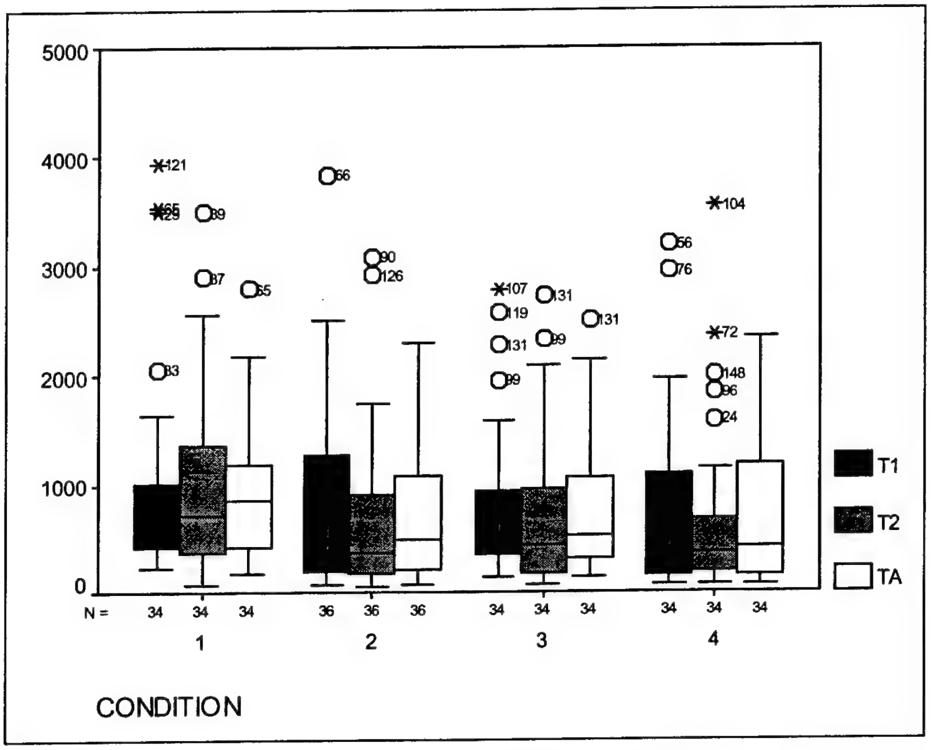


Figure 13 - Raw Score Boxplots by Group

As can be seen in this view for data depiction, the scores produced for the participants for Trial 1, Trial 2, and the two-trial average had large ranges, coupled with their large standard deviations (some even larger than their means), a method to compress the data was searched for that could effectively convert the data in hopes of reducing the large size of the standard deviations and reducing the number of outliers. Therefore, transformations of the data that were attempted included square/cube root conversions and logarithmic

conversions. Using a base 10 logarithmic conversion of the scores proved to provide the best compression of the data and elimination of outliers, while at the same time, maintaining the integrity of how the data relates to each other among the various treatment groups. Figure 14 (below) demonstrates the improvements made by converting the scores to their base 10 logarithmic equivalents.

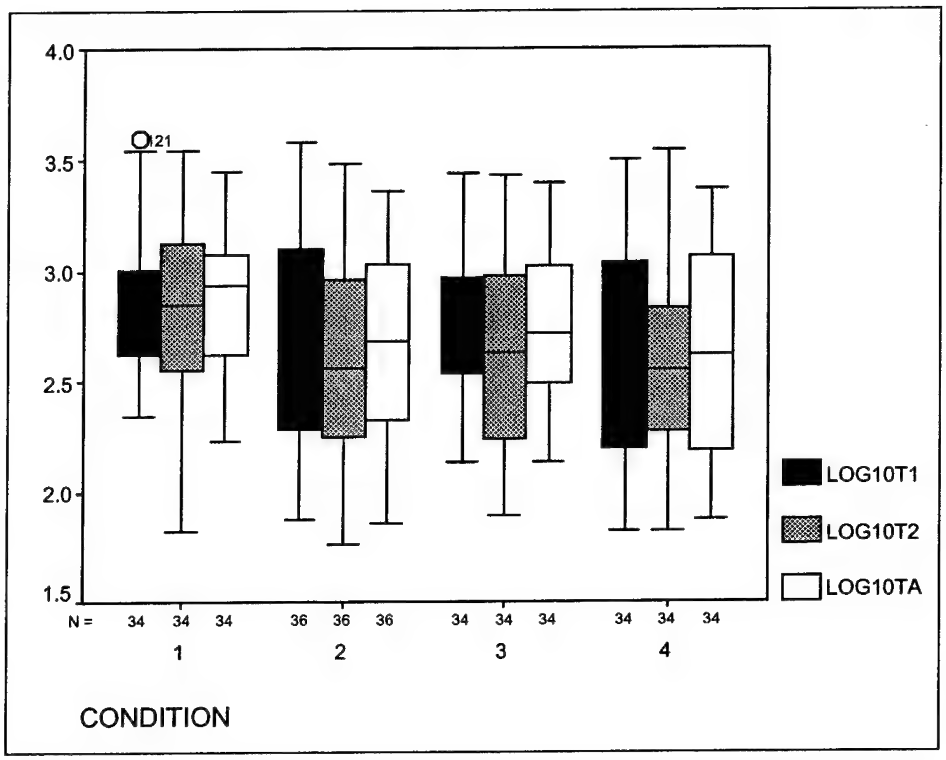


Figure 14 - Base 10 Logarithmic Transformation of Scores by Group

Data Description

Descriptive information for all variables (a total of 54 data points were collected for each participant) is not shown in this section. Only pertinent variables that may have some bearing on the research are discussed (for a complete list of all variables collected in the research, refer to Appendix I). Descriptive information for pertinent variables, along with their SPSS codes, is as follows:

1. Gender: Male = 1, Female = 2
2. Age: Age in whole years
3. Grad: SUNY status: 1 = Undergraduate student, 2 = Graduate student
4. Exp: Years professional experience
5. TT: Total time used to complete the experiment
6. Log10T1[††]: Base 10 logarithmic conversion of the T1 score
7. Log10T2[††]: Base 10 logarithmic conversion of the T2 score
8. Log10TA[††]: Base 10 logarithmic conversion of the TA score
9. Delta[††]: Change in scores between trials (Log10T1 – Log10T2)
10. TS: Test score (knowledge survey result)
11. SA3: Self-assessment survey question 3 (1 to 5 scale – 1 represents strongly disagrees, 5 represents strongly agrees)
12. SA3FIVE[††]: Identifies participants that scored SA3 with a "5"

Motivation Factor

At this point, special emphasis needs to be made regarding how the level of effort was operationalized during the data analysis process. Recapping the initiative in this area, Hsiao (1999) discovered only three methods of measuring "level of effort" on behalf of participants in a dynamic decision-making (DDM) study. They are: First, is the amount of decision time (how long does it take to make a decision). Second, is the amount of information use for specific information items (is the participant using the information

[††] These variables were not "collected," rather, they were computed within SPSS.

provided in the experiment). Third, is the amount of discussion among participants (do they seek each other's help when allowed by the experiment).

This researcher, interested in this aspect of DDM, posits that if human subjects really tried hard, they would perform well with respect to the various treatments they are exposed to in the current experiment. The idea was to administer a post-experiment self-assessment survey where subjects would be able to self-identify: 1) how hard they were trying, 2) their knowledge of the game, and 3) their interest in the research project.

After several analyses, it was discovered from the subjects' self-assessment survey that their "task knowledge" and/or their "interest in the research" were not good predictors of their effort. However, the statements regarding their performance in the self-assessment survey may have been somewhat ambiguous – except for one statement. The variable, SA3 (self-assessment survey item #3), stated: "I did my best in performing during this experiment;" the position of this statement is establishes that if someone was really trying hard, he or she would give this a top rating of "5" (meaning that they strongly agree with the statement). It is believed that this one measure alone can identify a subject who was "motivated." All others ranking this statement less than "5" is considered to be unmotivated, or at least, not as motivated as the researcher would like them to be.

Extending the logic of motivated vs. unmotivated, the boxplots below in Figure 15 show a marked difference from the boxplots shown earlier for all cases. Here, the motivated individuals by group have been separated from those who are unmotivated. Clearly, from a descriptive point of view, the differences in performance between those

who are motivated and those who are not appear to be noteworthy, and warrant further investigation.

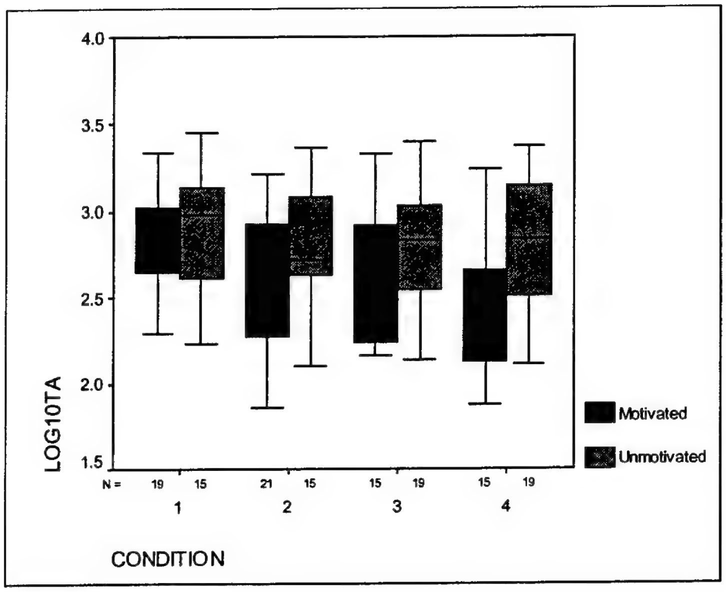


Figure 15 - Performance Comparisons s of Motivateds vs. with Unmotivateds

Limits of Research Design

The greatest threat to the research design is that of being able to provide alternative explanations of the results. Where evidence was found in support of the research hypotheses, the researcher must carefully consider what other explanations there would be as to why the results turned out the way they did (and possibly that the results were not a consequence of the experimental intervention).

Where the research indicates a positive relationship between the intervention and the overall participant performance, could this be explained by one, or some, of the kinds of confounds suggested by Bernard (2000). They are:

1. History
2. Maturation

3. Testing
4. Instrumentation
5. Regression to the mean
6. Selection of participants
7. Mortality
8. Diffusion of treatments

For the current experiment, no threats can be discerned that are caused by history, maturation, instrumentation, regression to the mean, mortality, and diffusion treatments. Conversely, testing is a real threat, since participants were given a practice trial along with two or three measured trials. It is possible that performance improves over time because the subject has merely gotten used to the simulation. It will be important to analyze any improvements made by the intervention groups with any improvements made by the control groups. The analysis of variance using the change in scores between the first and second trials was used to ferret out testing, or iteration, effects.

The selection of participants is most likely the strongest threat to external validity under the current research design. Although the researcher randomly assigned participants to various treatments, one must realize that a convenience sample may still affect the external validity of the research design. It is possible that if the same experiment was to be conducted with a different sample population (e.g., government bureaucrats, business leaders, store clerks, laborers, etc.) the results may be different. This is even more of a factor where the research provides evidence in support of the hypotheses. In this event,

future studies are recommended to see if replication of results can be found among other or, more disparate subjects.

In the final analysis, the researcher believes that the current experimental design has a high level of internal validity. However, there exists the liability of low external validity. It will be very difficult to generalize the results of a very controlled lab experiment to decision makers in the real world. This is compounded by the fact that the experiment contains a high level of artificiality. Yet, where the experiment provides evidence to support the hypotheses, it represents a stepping-stone for future research to take a more empirical approach to the overall questions raised in the study because it shows that decisions within complex systems can be improved with proper training, knowledge, motivation, and proper focus on pertinent cues. Therefore, the results will allow for empirical interventions in natural settings whereby the decision makers could have their own work environments modeled, and then be trained on which cues (along with judgment functions) to apply to their everyday problem solving.

Limitations related to the experiment surveys are most closely related to the fact that the written surveys were not the primary method of data collection. It was ancillary to the computer simulation experimentation process. The written surveys were used to satisfy the researcher that the subject: a) was interested in the research, b) understood the tasks, and c) participated fully (Appendix H), and to determine their knowledge of the dynamics of the microeconomy (Appendix G).

Internal validity is defined by the degree to which one can be certain that changes in the dependent variable are caused by the treatment (Bernard, 2000), and that the variables

are linked together in a relationship (Krathwohl, 1998). Here, the researcher is attempting to establish such a relationship between participant self-perceptions and participant performance. Granted, it may be possible to associate poor self-assessments with poor simulation performance scores; however, one must consider the possibility of associating high self-assessments with superior performance. In other words, subjects who are trying hard should produce better game scores.

External validity would come under scrutiny for the following: That the self-assessment survey indicates that participants did not give it their best effort, or they did not understand the tasks, or were simply not interested. It is the opinion of the researcher that to make any generalizations of performance when confronted with low self-assessments would be meaningless.

Conversely, the external validity of the study is enhanced where participants are shown to have done their best from a self-assessment perspective. The rationale for this inference is that the generalization of the treatment will not be able to be discounted from the position that the subjects failed to provide adequate participation. In other words, the final simulation performance results will be more important when attributed to those participants who try their best in performing.

Ethical Considerations

Anytime one uses human subjects in a research design, ardent ethical considerations must follow. The biggest item the researcher attempted to maintain was to treat subjects with respect. Not all performed brilliantly; some even performed poorly, yet that was

expected. However, they were all treated the same – with dignity and appreciation for having volunteered their time and effort.

The issues of confidentiality and privacy should also not go without mention. It was critical for the human subjects to know that their performance in the experiment, and their answers and comments provided in the surveys, were completely private and confidential. Only two people will ever know the individual's experimental performance: the researcher and the participant. To break the trust established by the researcher and human subject would decrease the quality and quantity of any future research. All endeavor has been made to maintain this trust between researcher and participant.

Chapter 5

FINDINGS

Descriptives

Considering the descriptives statistics for all participants (Table 2 below), the gender difference is near evenly split (53% female). The age of participants ranged from 19 to 51, but the mean and standard deviation indicate the age spread was predominately young. A similar skew occurs with experience – low experience. Graduate students made up 34% of participants tested. The total time participants took to undergo the experiment ranged from 55 to 173 minutes (average time was 94 minutes). Test score data (results of the knowledge survey) ranged from 26 to 91 with a mean of 55 and was evenly distributed (see Figure 16). Regarding the Base 10 Logarithmic scores obtained, the three variables measured have standard deviations that are very small compared to their means. The Delta (change from the Trial 1 to Trial 2 scores) ranged from a -.76 (participants doing worse in the second trial) to 1.31 (participants doing better in the second trial) and had a mean of .07 (indicating that the number of participants who did better in the second round of play was larger than those in the first). The final two variables, SA3 and SA3Five simply show the range of motivation (SA3) and that the number of motivated participants (SA3Five) represented 51% of the sample population.

Descriptive Statistics of Pertinent Variables[a]

	Minimum	Maximum	Mean	Std. Deviation
Gender	1	2	1.53	.50
Age	19	51	23.43	4.81
Graduate Level	1	2	1.34	.48
Experience	0	33	1.71	4.10
Total Time	55	173	93.55	21.72
Test Score	26	91	54.75	15.09
Log10T1	1.83	3.60	2.73	.40
Log10T2	1.76	3.55	2.66	.45
Log10TA	1.86	3.45	2.74	.39
Delta	-.76	1.31	.07	.41
SA3	2	5	4.38	.74
SA3Five	0	1	.51	.50

a. N = 138

Table 2 - Descriptives for All Participants

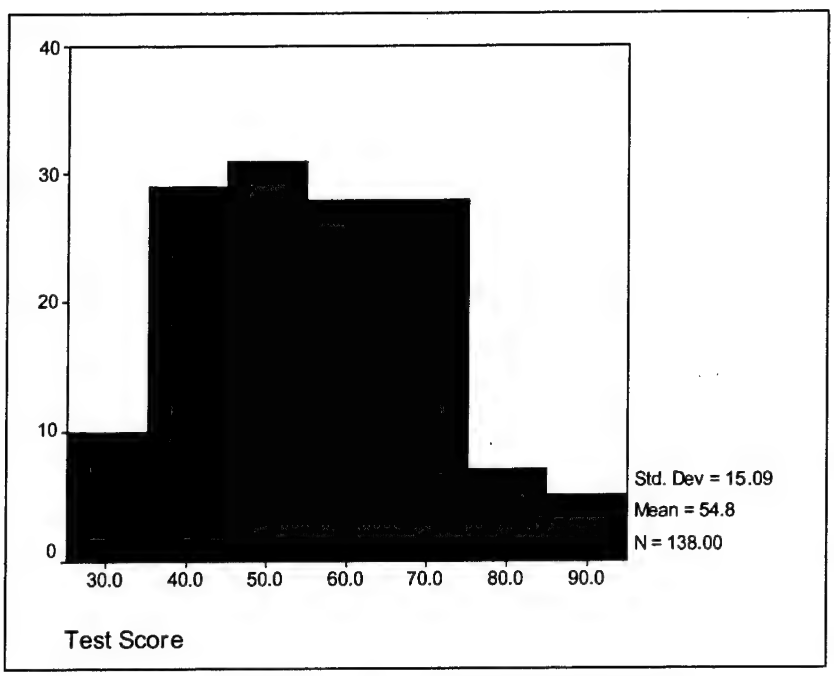


Figure 16 - Histogram of Test Scores (Knowledge Surveys)

It was important, however, to determine if the participants were randomly distributed among the four conditions established by the method of study (Bois instruction vs. no

instruction, and Richardson and Rohrbaugh rule vs. no rule). To do so, a one-way analysis of variance was conducted for each of the variables in Table 2 (cross-checked against each of the four research conditions) to see if any non-random assignments could be found as significant ($p < .05$). The result of this test indicated that no variable was found to have a significant non-random assignment. This means that the assignment of participants to the various treatments was indeed statistically random.

Research Hypotheses

To facilitate reader comprehension of the analyses in the remainder of this chapter, a review of the research hypotheses are presented. Assuming there are ways to improve human performance in the face of time-delayed feedback dynamics, the following hypotheses were projected for this research thesis:

1. If information and knowledge about a system are better understood, participant performance will improve.
2. If participants are provided with a decision rule that focuses their attention on proper cues and how to weigh their importance, their performance will improve.
3. Participants reporting greater effort during the experiment simulation will out-perform those who do not.

Analysis of Variance

As a reminder to the reader, Figure 17 (below) is provided in order to show the specific treatments and their associated conditions.

Conditions and Treatments	No Rule	Receives R & R Rule
No Bois Instruction	I	II
Receives Bois Instruction	III	IV

Figure 17 - Human Subject Random Group Assignments

The first hypothesis, improving knowledge and information, is represented by applying the Bois instruction. The second hypothesis, focusing participant attention on proper decision cues and weights, is identified by the application of the Richardson and Rohrbaugh Rule. The last hypothesis, participants reporting a greater level of effort, is not directly reflected in Figure 17, however, it was included as a third factor in the analysis of variance.

The main effects observed in the analysis of variance are shown in Figures 18 through 21 (below). Four ANOVAs were performed. They included analyses of the first trial, second trial, the two-trial average, and a delta (Trial 1 minus Trial 2). The predominant trend that is seen in the following figures is that the mean performance scores for the first and second trials, along with the two-trial average, show improvement when either the Bois instruction, or Richardson and Rohrbaugh rule, is applied. Additionally, there is a pronounced improvement in scores on behalf of participants who were assessed as motivated over those who were not.

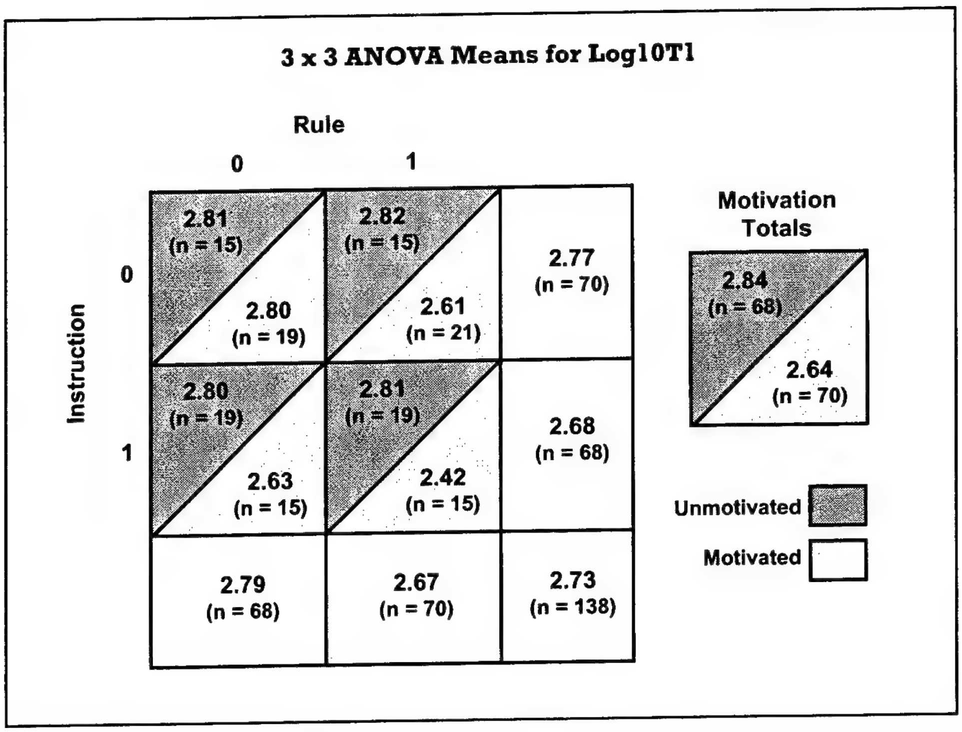


Figure 18 - ANOVA Findings for Trial 1

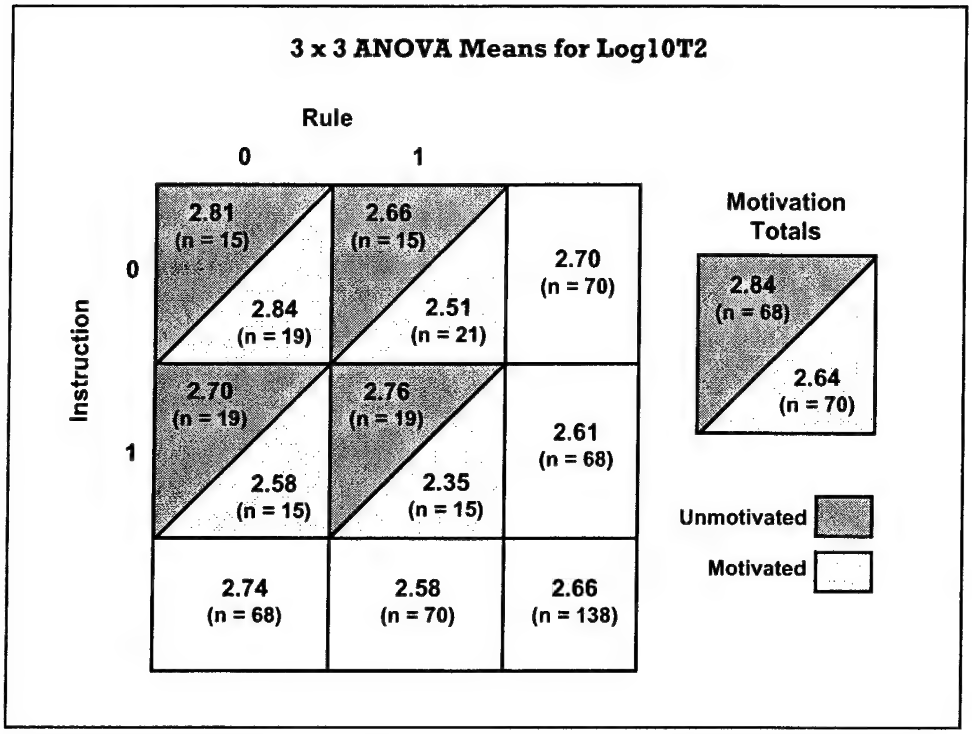


Figure 19 – ANOVA Findings for Trial 2

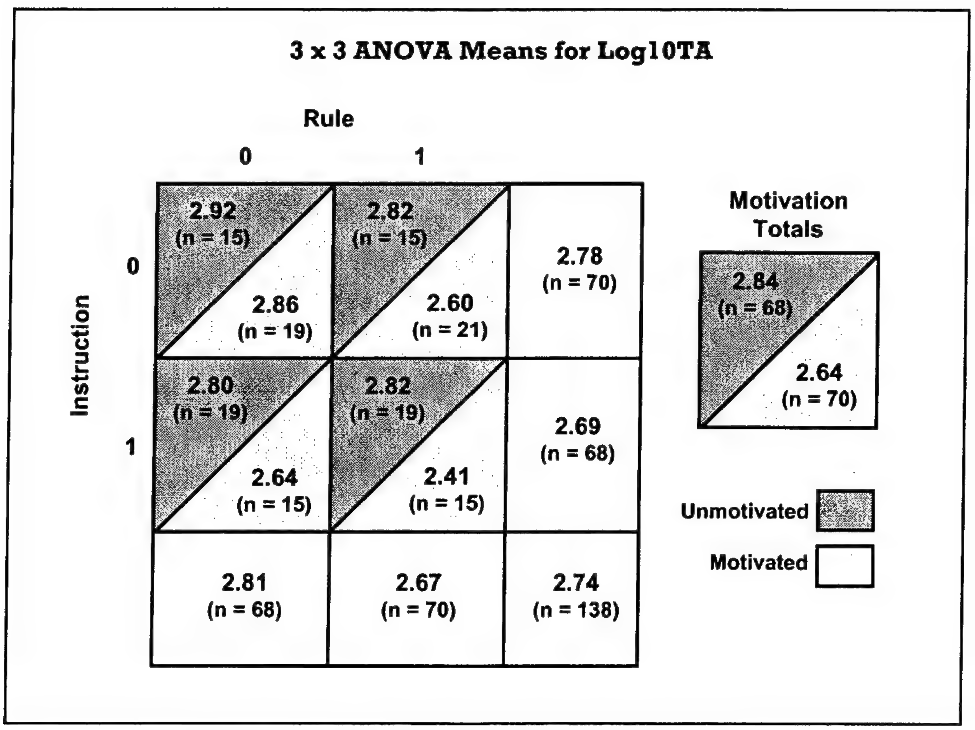


Figure 20 - ANOVA Findings for the Two-Trial Average

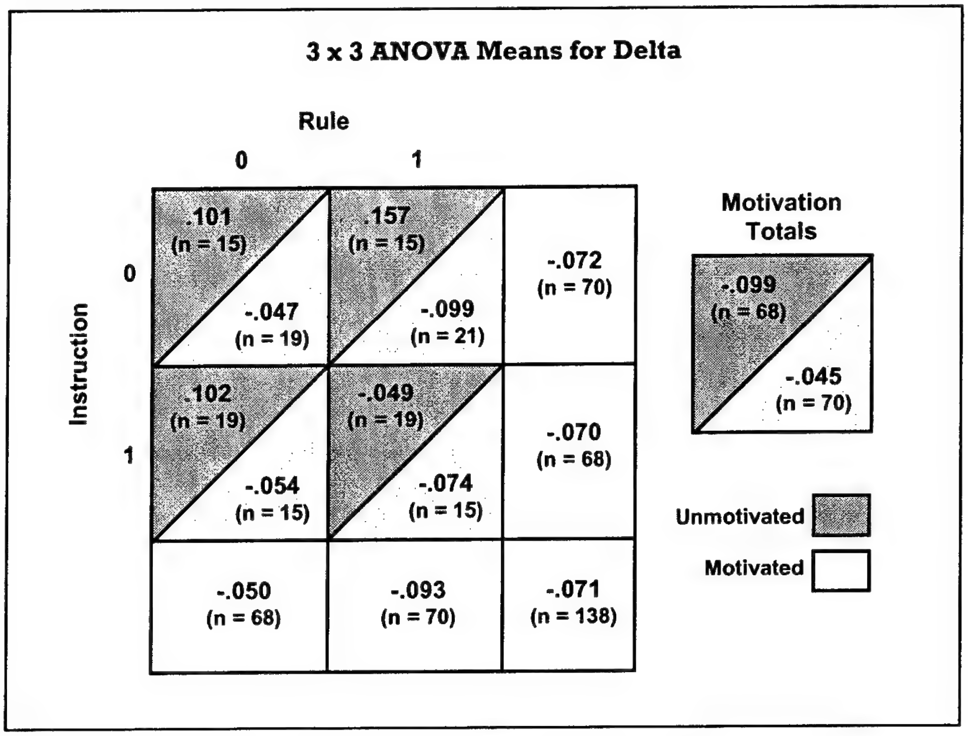


Figure 21 - ANOVA Findings for Two-Trial Delta

The analysis in the Delta category (change in performance from Trial 1 to Trial 2) yielded no worthy information – no findings were found to be significant. This is probably due to having 57 cases performing worse in the second trial.

In an attempt to better show (graphically) the results of the four ANOVAs (Trial 1, Trial 2, the Two-Trial Average, and the Delta between trials), the following Figures 22 through 25, demonstrate the results of each of the four analyses. What is important to remember is that the circles represent participants not receiving the Bois instruction, the triangles represent the reception of the Bois instruction, and the left aligned circles and triangles represent participants not receiving the Richardson and Rohrbaugh rule (compared to those aligned on the right side of the chart – they received the rule). Motivation is also separated by color as indicated.

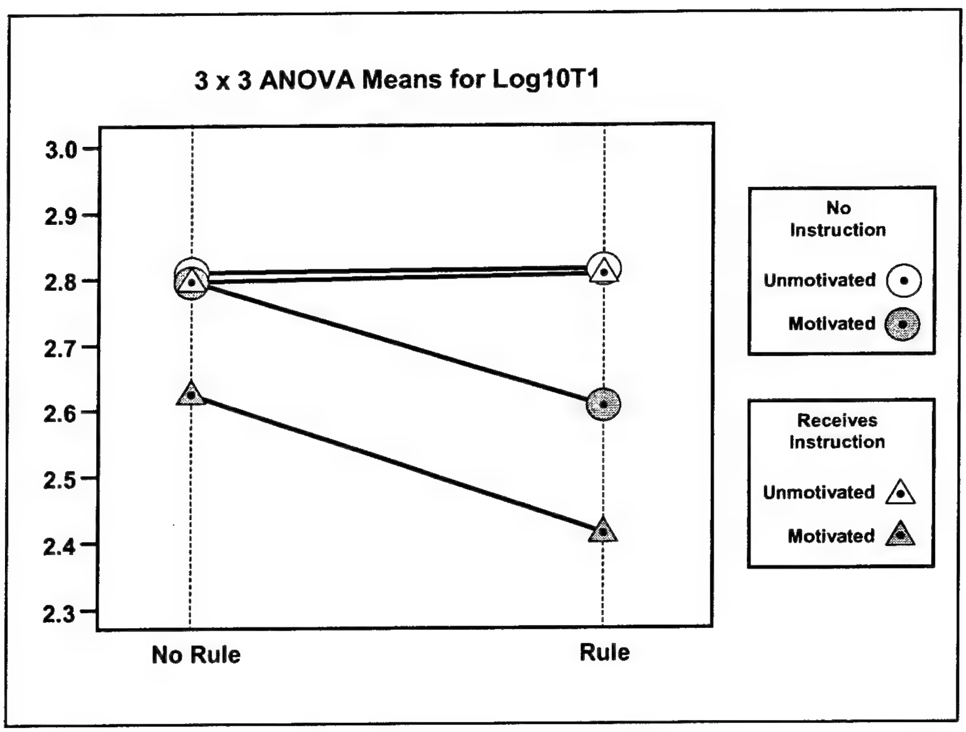


Figure 22 –Graph for Trial 1 ANOVA

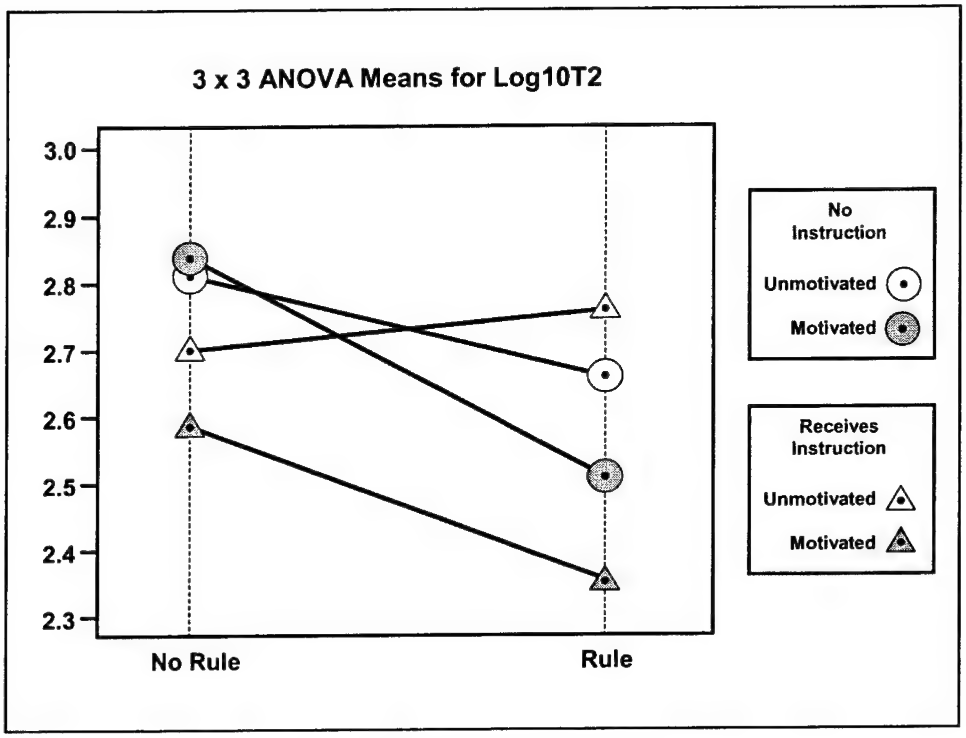


Figure 23 –Graph for Trial 2 ANOVA

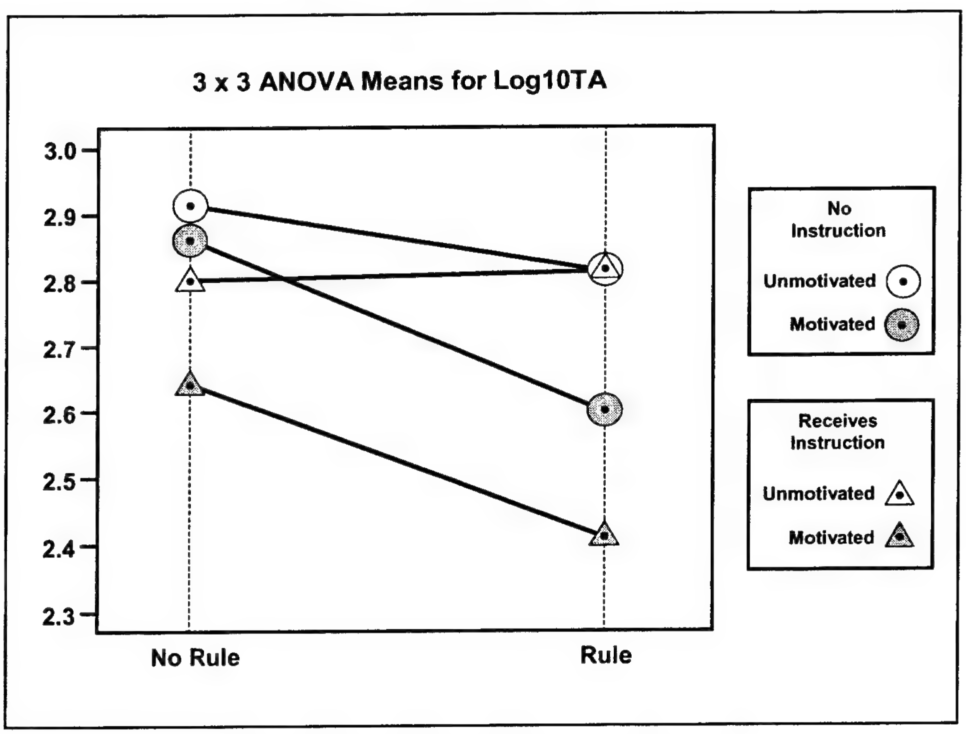


Figure 24 - Graph for Two-Trial Average ANOVA

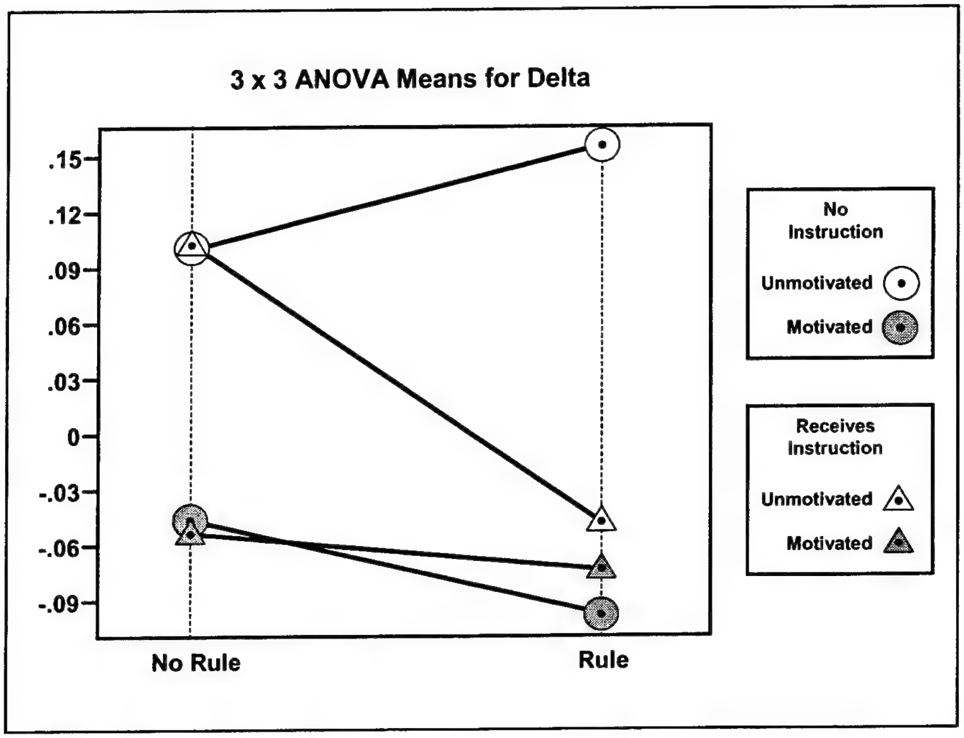


Figure 25 - Graph for Two-Trial Delta ANOVA

What is important to discern in Figures 22 through 25 is the difference in performance between the motivated/unmotivated subjects. For example, in Figures 22, 23, and 24, the unmotivated subjects show little or no improvement between those who received the Bois instruction and those who did not. The same relationships can be discerned between those subjects receiving the Richardson and Rohrbaugh rule to those who did not. However, what is critically important to observe, is that among motivated subjects, the differences in performance between the those who received the Bois instruction and those who did not, along with the comparison of subjects receiving, or not receiving, the Richardson and Rohrbaugh rule, all perform as hypothetically predicted (this precludes observations in the Delta category, which are statistically insignificant).

The following table shows the F-ratios obtained in the analysis of variance. Although the instruction, rule, and motivation factors had no significant bearing on the Delta variable, all three main effects were significant when the two-trial average was used as the dependent variable. Motivation also had a significant main effect in the first and second trails. Additionally, the rule had a significant main effect in the second trial.

F-Ratio of Main Effects

	Log10T1	Log10T2	Log10TA	Delta
Instruction	3.11	2.11	4.13*	.01
Rule	3.31	4.72*	5.05*	.35
Motivation	11.07**	4.74*	10.93**	.64

* Sig. at the .05 level
** Sig. at the .001 level

Table 3 - F-Ratios of Main Effects for Instruction, Rule, and Motivation

Table 4 shows the F-ratios of the interactions between the main effects of the Bois instruction, the Richardson and Rohrbaugh rule, and the motivation factors. All F-ratios were found to be non-significant.

F-Ratio of Main Effect Interactions

	Log10T1	Log10T2	Log10TA	Delta
Instruct * Rule	.09	1.08	.31	.64
Instruct * Motivation	.80	1.96	1.25	.67
Rule * Motivation	1.41	2.55	2.43	.40
Instruct * Rule * Motivation	.22	.13	.11	.33

Table 4 - Interaction F-Ratios

Motivation Factor Explained

In the third hypothesis, "participants reporting greater effort will out-perform those who do not," two two-way analyses of variance were performed to determine the significance of the Bois instruction and the Richardson and Rohrbaugh decision rule with those who are motivated, and those who are not. Table 5 shows the results of these two analyses.

Log10TA Motivation F-Ratios

	Unmotivated	Motivated
Instruction	.44	4.80*
Rule	.25	7.02**

* Sig. at the .05 level
** Sig. at the .01 level

Table 5 - Motivation F-Ratios\

As a final addendum to this section, another very interesting discovery was made when comparing participant knowledge survey scores to their self-assessed motivation levels. The boxplots below in Figure 26 show a very different level of performance in test scores (knowledge survey) between the two motivation levels. Below the boxplots (Table 6) are the descriptives for these two levels of measurement, along with a two-tailed significance test of their means. The means differences between them are not only large (10 points), but their significance is at the .0001 level.

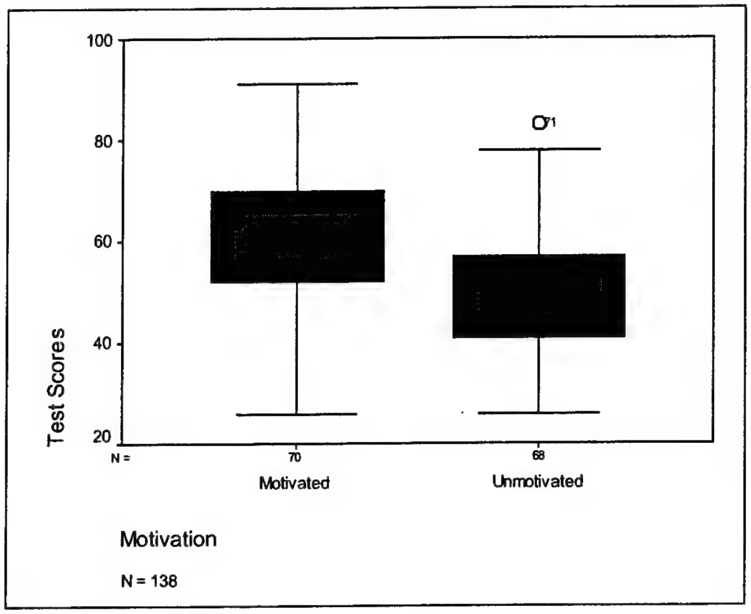


Figure 26 - Test Scores by Motivation

Motivation Comparisions

	Mean	Std. Deviation
Motivated	59.61	15.00
Unmotivated	49.75	13.55
Mean Difference	9.86****	

****Sig. at the .0001 level

Table 6 - Two-Tailed T-Test of Motivation between Groups

The significant differences in test scores may attribute, in some way, to the increased performance of simulation scores between the two motivation levels of subjects.

Anecdotal Observations

Given the statistical findings (above), other observations about the experiment must be highlighted. For example, although it is not quantified, participants who received the Richardson and Rohrbaugh decision rule used it in different ways. The researcher observed

that when given the rule card, participants often times would simply discard it. At other times, they would try to perform the calculations prescribed by the card, only to abandon the rule card over time. Yet, others would follow the prescriptions of the rule to the very end of the experiment.

The "score" in the simulation was also another area of concern. It seemed that several participants would focus too much attention on this output of the game interface. For example, several participants would preoccupy themselves with trying to obtain a lower score versus trying to properly balance supply and demand.

Depreciation did not seem to be fully understood by many participants. It is the ONLY means of reducing capital stock. In other words, when current capacity was too large for the desired production, there were several subjects who neglected to simply order "zero," in order to lower their capital stocks. Many participants failed to appreciate that during times of excess capacity, depreciation could be used to assist them in lowering their production capabilities in order to try and balance their supply with demand.

Finally, as an overall observation, many participants found the simulation very difficult to understand. This is from an observational point of view and could not be corroborated with self-assessment data.

Ambiguities

When considering the questions posed during the self-assessment survey portion of the experiment, it was discovered that those statements dealing with "task knowledge" and "research interest" where not of any statistical value. However, statements dealing with

self-assessment of "performance," several ambiguities were discovered that may have led participants to misunderstanding what exactly was being presented. For example, self-assessment statement #7 says, "When provided with a set of decision cues to follow, I followed them all the time." The problem with this statement is the word "cues." What does it mean? Is it likely that the average participant would not understand what is being stated. Additionally, this statement was geared toward subjects who had received the Richardson and Rohrbaugh decision rule, and/or subjects who received the Bois instruction. These participants were pointed to specific elements of the simulation and how to react to them; however, they were never told that these elements were "cues." This can lead to very inappropriate understanding of the statement. Self-assessment statements #8 and #12 were found to have similar ambiguities. Only statement #3 was discerned to be unambiguous.

The only other item that can be considered ambiguous deals with the verbiage of the Richardson and Rohrbaugh decision rule card that was used for participants receiving the rule. The card (see Figure 11 on pg 56) has two sides. On the first side, the participant is told that they have hired a reputable consultant to assist in STRATEGEM-2 decision making. The participant is reminded that they must remain diligent with using the decision formula presented by the consultant (which is the Richardson and Rohrbaugh decision rule). A sample "work through" of the rule is also presented on the front side of the card. On the reverse side of the card is a layout of the formula that the participant can use by simply "plugging in" numbers found on the game interface. The layout then provides a

step-by-step process whereby the participant then arrives at a calculated game input – a number to be used for capital goods orders.

Several ambiguities were discovered after the fact that has led the researcher to wonder how effective the treatment of the Richardson and Rohrbaugh decision rule was upon game play. For example, on the front and reverse side of the card, the term "shortfall" was clarified for the participant. Directly below this statement was added verbiage stating: "*If this figure computes to less than zero, use zero.*" An ambiguity occurs because this added statement was meant to relate to the computed final "total orders" produced by the Richardson and Rohrbaugh decision rule and not to the "shortfall" amount. Additionally, more ambiguity occurs because the rule does not address when computations end with a value that is not evenly divisible by 10 (because the game interface rounds all values to their nearest 10).

The final ambiguity discovered was on the reverse side of the card whereby the shortfall amount was shown to be added to the computed depreciation value. This is correct; however, if the shortfall computes to a negative number, the participant needs to know that instead of adding, they would now be subtracting the shortfall amount from the computed depreciation value. Appendix J contains an improved Richardson and Rohrbaugh decision rule card for any future research desiring to use this approach in the STRATEGEM-2 game.

Given these findings regarding ambiguity with the Richardson and Rohrbaugh decision rule, the researcher was uncertain as to what their effects are on the results of those treatment groups that were exposed to the rule. The reason being is that in the face of the

ambiguities, several participants were able to use the rule card and achieved very low scores. Others did not, but was that a result of the ambiguities, or that possibly they simply discarded the rule (as was observed by the researcher as an anecdotal finding), or was it that they were simply not analytically inclined to fathom the directions proposed by the Richardson and Rohrbaugh (1990) formula? These questions cannot be fully resolved. However, as a minimum, mean scores of the treatment groups using the Richardson and Rohrbaugh decision rule were at a level consistent with hypothetical predictions (regardless of their statistical significance). Findings for these data, therefore, will remain as stated. As a final statement for this specific ambiguity, it is felt that if an improved Richardson and Rohrbaugh rule card were used (Appendix J), it might result in better scores for those participants exposed to the rule.

Chapter 6

CONCLUSIONS

First Hypothesis: The Impact of Knowledge and Information

The first hypothesis in the research postulated: If information and knowledge about a system are better understood, participant performance will improve. The control for this hypothesis was represented by participants not receiving the Bois instruction. The treatment was to introduce the Bois instruction to another set of randomly assigned subjects. The research question associated with this hypothesis asks: Can proper/adequate knowledge and information about the system be taught to participants? The significant F-ratios found for the mean average two-trial performance suggest that this may be so. However, caution must be exercised. For example, were these improved performance scores due to iteration? Cognitive style? Or, participant learning style? The answers to these questions are not known from the current study as these areas of interest were not measured during the experiment.

Second Hypothesis: The Impact of Decision Support

The second hypothesis of the study states: If participants were provided with a decision rule that focuses their attention on proper cues and how to weigh their importance, their performance will improve. The control for this hypothesis was represented by participants not receiving the Richardson and Rohrbaugh rule. The treatment was to introduce the Richardson and Rohrbaugh rule to another set of randomly assigned subjects. The research question associated with this hypothesis asks: Can participant performance be improved

via decision cues and weights? As in the first hypothesis, the significant F-ratio scores for the two-trial average suggest that improvements to the decision-making process can be made through the use of cues and weights. Again, did iteration, cognitive style, or learning style have a factor in this finding? The answer to this question cannot be determined from the current research design.

Third Hypothesis: The Impact of Level of Effort

The final hypothesis of the study suggested: Participants reporting greater effort during the experiment simulation will out-perform those who do not. This hypothesis is used in an attempt to answer the following research question: Can a participant's self-assessment of level of effort be used to better determine their own experiment performance? Level of effort was operationalized through motivational self-assessment on behalf of the participant. The discoveries made in this area were found to have a noteworthy impact upon experiment results.

The first of these discoveries was found in the significant F-ratios for each trial (and the two-trial average) of the experiment. These ratios suggest that subjects who really try hard to implement experiment interventions consistently have a greater effect (improved performance) than those who do not. This is an important finding in light of the corpus of the dynamic decision-making literature for it posits the question of how important other research findings have been because they have not been filtered/differentiated for motivation factors.

The angle of this specific portion of the research is to determine if there is a masking of the results obtained that can somehow be peeled away, revealing a better understanding of

the experiment treatments. Specifically, when the data set was divided between motivated and unmotivated participants (as self-assessed from the viewpoint that "they did their best" while participating in the experiment), it was found that those who self-assessed themselves to be motivated, outperformed those lacking full motivation.

For the motivated subjects, significant F-ratio results were found for the motivated participants versus the unmotivated participants. Therefore, it is possible that lower motivation levels (those that are not fully motivated) mask the intended treatments that are designed to improve decision making in the STRATEGEM-2 environment.

Using the motivation discriminator reveals another interesting facet of the research. Test scores (results of the knowledge survey) averaged about 55 percent (on a 0 to 100 percentage scale) for all participants. Yet, when considered independently between those subjects that were motivated and those that were not, the mean scores were about 60 percent for motivateds versus 50 percent for unmotivateds. This was a clear indication that the motivational level produces improved results upon performance, and it was found significant at the .0001 level.

Discussion

This dissertation project began with the notion that the Sterman (1989a) experiment with STRATEGEM-2 may have been flawed with respect to the misperception of feedback hypothesis. Specifically, participants in the simulation performed poorly in light of having "perfect knowledge and perfect information" while undergoing the rigors of play.

It is the current research initiative that the Sterman (1989a) observations regarding the misperception of feedback hypothesis remain accurate to some degree. This means that participants perform poorly because they fail to properly perceive the time delays in the system, and they fail to understand the effect of their decisions to their environment. These elements of the misperception of feedback hypothesis cannot be eliminated from current findings, however, what cannot be corroborated, is the perfect knowledge and information premise made by Sterman (1989a). For example, as was performed in the Howie and others (2000), knowledge of the simulation and system environment was tested in the current study. The results in this portion of the experiment once again clearly demonstrate that participants do not possess perfect knowledge of the system. Regarding whether participants possess perfect information is also debatable. Although Howie and others (2000) were able to demonstrate how an improved simulation interface works toward improving the information about the system to the participant and, in turn, contributes toward better participant performance, one cannot say that the information presented is perfect. This facet was not tested by Howie and others (2000), or by Sterman (1989a), yet was claimed to exist by Sterman (1989a). The current study does not profess that such an ideal of "perfect information" exists, and it cannot be determined how such a concept can even be measured.

The current research argues that the notion of perfect knowledge and information should no longer be a part of the misperception of feedback hypothesis. Rather, the opposite is more probable, that perfect knowledge and information are not a benefit enjoyed by participants.

Given that test subjects do not have perfect knowledge and information, it remains to know if they can be *taught* to make better decisions (the Bois instruction), or can they be *shown* to make better decisions (the Richardson and Rohrbaugh, decision rule). It is felt that this occurred on both counts – particularly when participants where screened for self-assessed motivation levels. However, caution is warranted; for it is unknown if the improvements observed from the interventions were not a result of other issues that were not measured (e.g. cognition, learning, and iteration). Given that significant effects in decision making were recorded for all two-trial average (Log10TA) scenarios, the results are still encouraging that either the Bois instruction, or Richardson and Rohrbaugh decision rule, were able to assist decision makers improve their performance over those subjects that lacked any assistance at all. The misperception of feedback hypothesis remains an important barrier towards effective decision making in dynamic environments; however, this study shows promise that decision makers can be aided in improving their decision-making skill in these environments.

The findings from the current research indicate three important factors that can be used to improve decision-making support in dynamic environments. The first factor is motivation. Clearly, this factor produced significant results across all levels of the simulation and it is important to take notice of it. Decision support researchers and consultants need to begin paying attention to this factor. Because the lack of motivation has a tendency to mask intended decision support interventions, it is imperative that decision support systems, particularly those in real world environments, consider ways to motivate decision makers to become motivated at a very high level. The methods to do so

are undetermined from the perspective of the current research. However, they could include such things as: monetary reward, enlisting decision makers to have a greater "stake in the outcome," and improved benefits (health coverage, retirement benefits, insurance coverage, compensation time-off, improved workspace, to name a few). This list is not exhaustive, yet, provides consideration for improving motivation among real-world decision makers who are operating in dynamic environments. For researchers, it represents a possible list of factors that can be used to determine the effectiveness of improving decision-maker motivation.

The second factor that can be used to improve decision-maker performance is instruction. The current research focused on increasing participant knowledge and interpreting information within a simulation environment. It is posited that the same can be translated to a real-world environment. Researchers and consultants in this area would need to focus more attention on trying to explain the dynamics of decision environments to decision makers. For example, in the current research, the Bois instruction focused very heavily on explaining the concept of "equilibrium" in the STRATEGEM-2 environment. This concept is not unique to STRATEGEM-2, but is applicable to most dynamic decision environments.

The third factor that can be used to improve decision-maker performance is rule based. The Richardson and Rohrbaugh rule not only had a significant effect on experiment results, it is possible that it provided great benefit to decision makers who had most difficulty in understanding the STRATEGEM-2 environment. It is opined that, possibly, decision makers who are most inclined to approach decision situations in an intuitive

(cognitive) manner would benefit most from such a decision rule. This is opposed to an analytically inclined decision maker who would rely more on his, or her, understanding of the environment to make better decisions. The effect of the Richardson and Rohrbaugh rule cannot be truly appreciated from the current research because cognitive faculties were not measured on behalf of the participants. However, a measurement of cognition among decision makers may possibly improve decision support interventions. This area is imperative for future research and should not be overlooked by real-world decision-support consultants.

The three factors identified as important for decision support within dynamic environments undoubtedly requires further study. The STRATEGEM-2 game is probably a very fine instrument to use to test varying hypotheses that will lead to better decision support in the real world. To do so, in the next chapter, a series of recommended studies is put forth. Particular attention is paid toward what the previous literature has covered, toward what the Bois research has posited, and toward what future research should include.

Summary

In summary, the current research concludes the following:

Perfect information and perfect knowledge should no longer be the premise of the misperception of feedback hypothesis, rather, improvements can be made with regard to decision makers operating in dynamic environments.

The intervention of a new instruction, one that teaches participants on the necessary knowledge to become a better decision maker within the dynamic environment, has shown to have had a significant effect ($p = .05$) towards improving decision-maker performance. This was observed from the main effect that the instruction had upon the two-trial average score.

The intervention of a decision support rule, one that directs participants toward specific cues and provides a weight for their importance, has shown to have significant effects ($p = .05$) toward improving decision maker performance. The rule's main effect was significant for the second trial score, as well as the two-trial average.

Participant self-assessed motivation had a significant effect ($p = .05$) on the second trial score, and it had very significant effect ($p = .001$) on the first trial and two-trial average scores. The motivation factor was also found to have a masking effect upon the results. This means that by measuring the subjects' motivation level, a truer overall picture of performance can be obtained. This was shown when a performance comparison was made between the motivated and unmotivated subjects. In this case, the instruction and the rule had large and significant F-ratios ($p = .05$ for the instruction, and $p = .01$ for the rule) for those participants who were motivated. For unmotivated participants, the instruction and the rule appeared to have no effect at all upon their performance.

Chapter 7

FUTURE RESEARCH

The results of this study are preliminary. Future research is required that will further add to the body of knowledge surrounding the current study's findings. Additional exploration will also be able to further address key issues that will better assist the development of improved decision support for decision makers within dynamic systems.

Future research using the STRATEGEM-2 simulation should address and/or possibly include the following:

The effects of iteration, cognitive styles, and learning styles should be explored in future studies. There exists a certain potential that these criteria may unveil more "masks," which may indicate a more accurate portrayal of research interventions.

The subject population requires expansion. The current research has relied (conveniently) on too small of a target population. Age, experience, education, background need to be expanded. The current pool of participants was very narrow in scope in these areas.

In the current research, the use of the STRATEGEM-2 interface is taught solely from a written perspective. It is possible that the interface can be "classroom taught" and would produce more uniform results and improve upon subject knowledge of the game dynamics.

The Bois instruction can certainly be improved upon. As a "first try" in producing an instruction via on-screen tutorial methods, a better computer-assisted instruction can certainly be devised.

Regarding the ambiguities discovered with the Richardson and Rohrbaugh decision rule card, using the improved card (Appendix J) is highly recommended.

The STRATEGEM-2 interface requires noteworthy changes. The first is the elimination of the game score being so prominently displayed in the center of the screen. Although it provides some sort of outcome feedback, it detracts players from staying focused on more important elements of the simulation. Second, the graphs showing historical feedback at the bottom of the interface could most likely be eliminated without any derogatory effects. Third, there needs to be a randomization of the goods sector demand within the simulation. Currently, the goods sector takes a single step increase of 50 units in year four during the game; using a randomization of increases or decreases would make for a more realistic environment and would also increase the validity of proposed decision support rubrics.

Finally, the greatest challenge of this type of research is to transform the dynamics of the experiment into real-world – natural – settings. Is it possible to find dynamic environments in the real world that can be used for discerning whether or not instructions or decision rules can be of added value within those dynamic environments? This is the crux of the current research that is hoped will someday bare important results.

Literature "Off-ramp" to Study

In the following, Figures 27 through 30, a review of the STRATEGEM-2 literature is presented once again. Additionally, various arrows point to sub-topics of dependent and independent variables; they recommend continued research, potential for improved research, and to elements not recommended for future research in a STRATEGEM-2 environment.

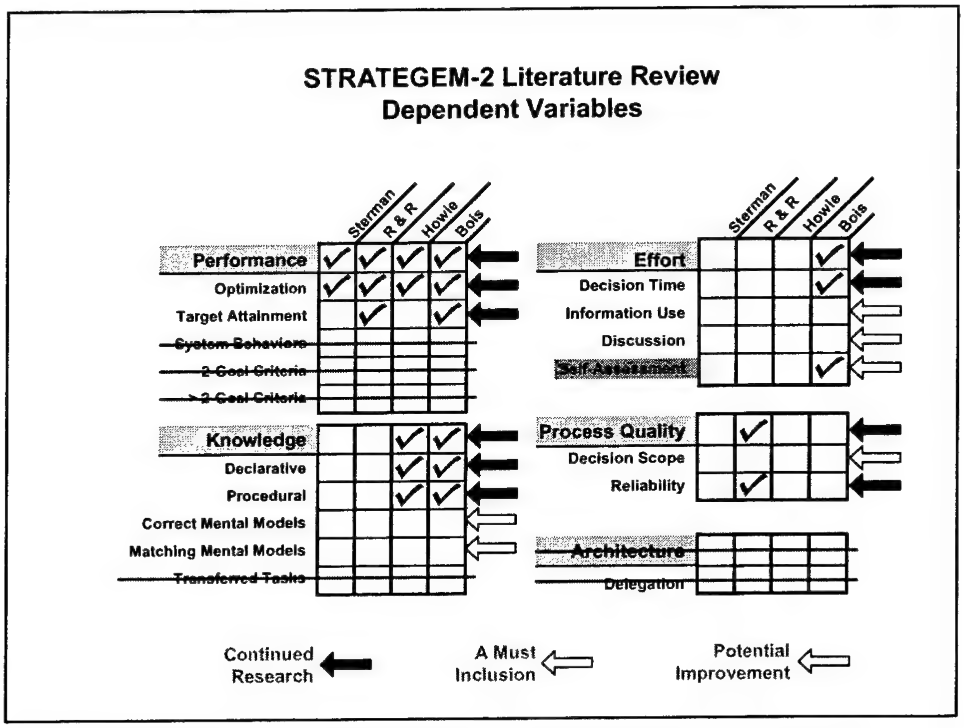


Figure 27 - STRATEGEM-2 Dependent Variable Summary Recommendations
*Self-Assessment has been added by the author

In Figure 27 (above), the author has placed emphasis on continued research in the area of self-assessed levels of effort. This area represents an important discovery in the current research and should not be overlooked in future experiments with human subjects using the STRATEGEM-2 simulation. Additionally, this area may also be of great importance to other studies/experiments in dynamic decision making.

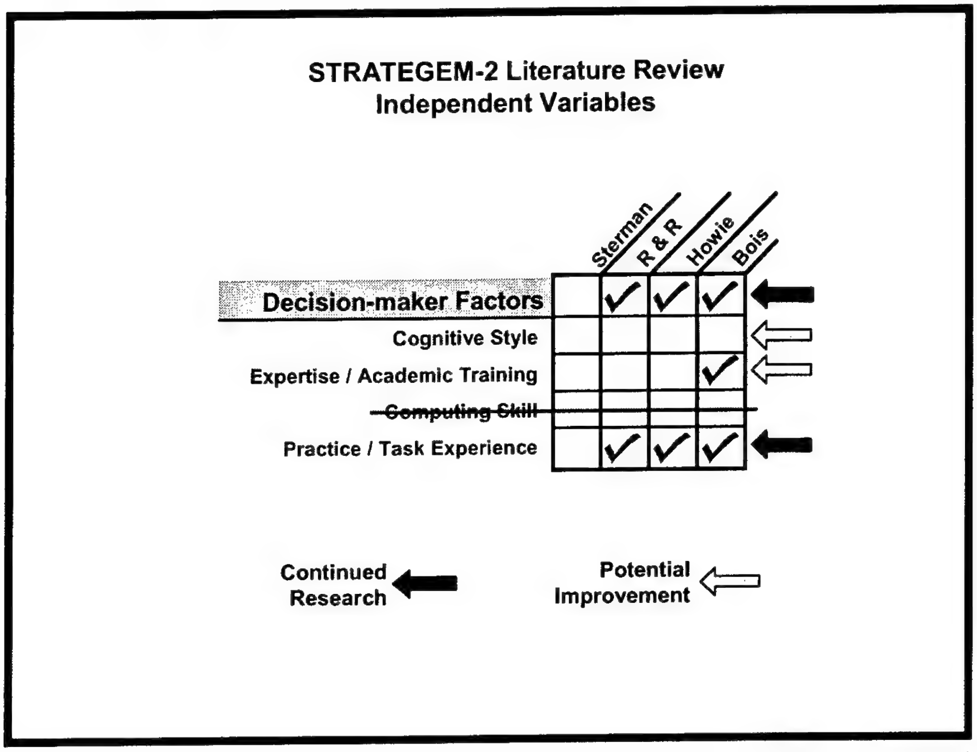


Figure 28 - STRATEGEM-2 Independent Variable Summary Recommendations(Decision-maker Factors)

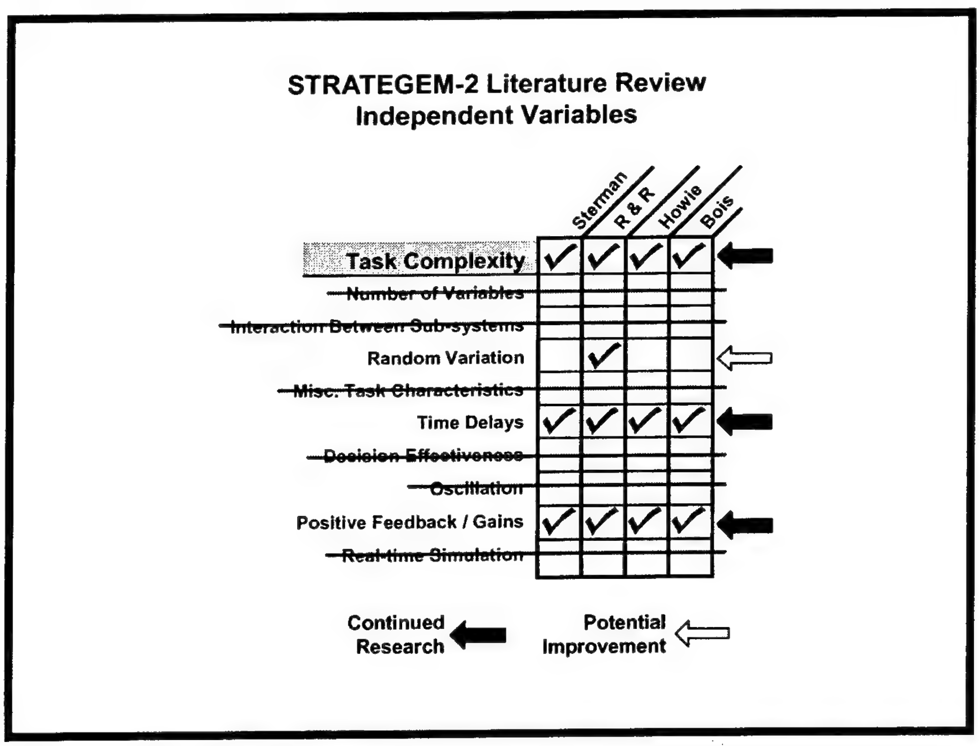


Figure 29 - STRATEGEM-2 Independent Variable Summary Recommendations(Task Complexity)

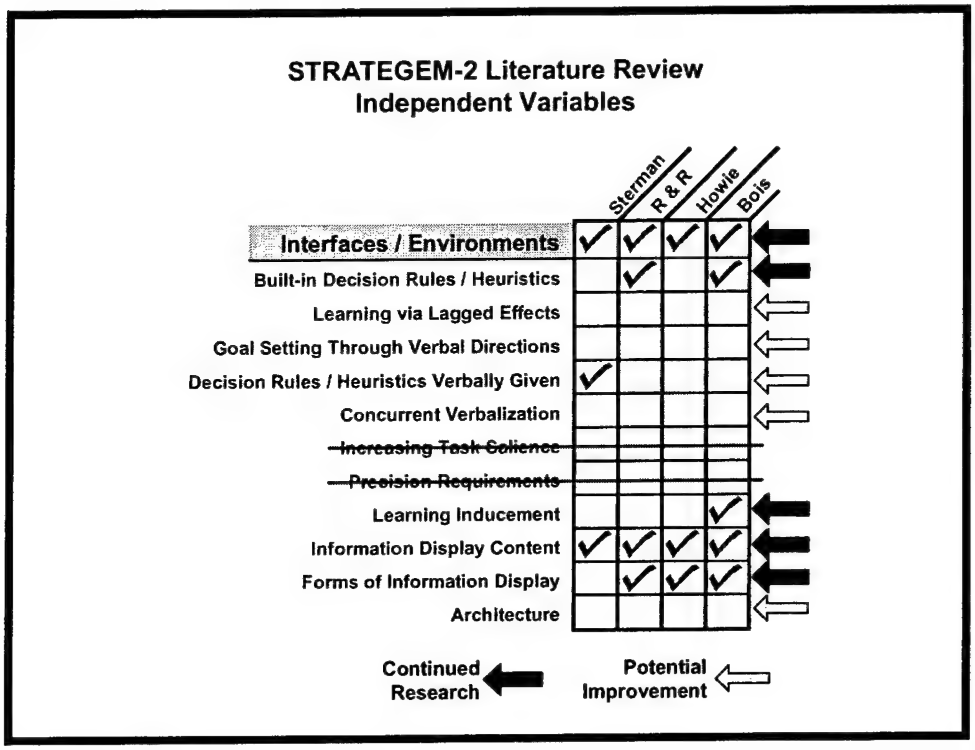


Figure 30 - STRATEGEM-2 Independent Variable Summary Recommendations(Interfaces / Environment)

REFERENCES

Anand, T., & Kahn, G. (1992). *Spotlight: A Data Explanation System.* Paper presented at the Eighth Conference on Artificial Intelligence Applications, Monteray, California.

Bakken, B. E. (1993). *Learning and Transfer of Understanding in Dynamic Decision Environments.* Unpublished Ph.D. Dissertation, Massachusetts Institute of Technology, Boston.

Balzer, W. K., Doherty, M. E., & O'Connor, J. (1989). Effect of Cognitive Feedback on Performance. *Psychological Bulletin, 106*(3), 410-433.

Bernard, H. R. (2000). *Social Research Methods: Qualitative and Quantitative Approaches*. Thousand Oaks, California: Sage Publications.

Berry, D. C., & Broadbent, D. E. (1984). On the Relationship between Task Performance and Associated Verbalized Knowledge. *The Quarterly Journal of Experimental Psychology, 36A*, 209-231.

Berry, D. C., & Broadbent, D. E. (1987). The Combination of Explicit and Implicit Learning Processes in Task Control. *Psychological Research, 49*, 7-15.

Berry, D. C., & Broadbent, D. E. (1988). Interactive Tasks and the Implicit-Explicit Distinction. *British Journal of Psychology, 79*, 251-272.

Brehmer, B. (1992). Dynamic decision making: human control of complex systems. *Acta Psychologica, 81*, 206-223.

Brehmer, B. (1995). Feedback Delays in Complex Dynamic Decision Tasks. In P. Frensch & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 103-130).

Brehmer, B., & Allard, R. (1991). Dynamic Decision Making: The Effects of Complexity and Feedback Delay. In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.), *Distributed Decision Making: Cognitive Models of Cooperative Work* (pp. 319-334).

Broadbent, B., & Aston, B. (1978). Human Control of a Simulated Economic System. *Ergonomics, 21*, 1035-1043.

Broadbent, D., FitzGerald, P., & Broadbent, M. (1986). Implicit and Explicit Knowledge in the Control of Complex Systems. *British Journal of Psychology, 77*, 33-50.

Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. Frensch & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 27-63). New Jersey: Lawrence Erlbaum Associates Publishers.

Collins, A., & Stevens, A. L. (1982). Goals and strategies of inquiry teachers. In R. Glaser (Ed.), *Advances in Instructional Psychology* (Vol. 2, pp. 65-119). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cooksey, R. (1996). *Judgment Analysis: Theory, Methods, and Applications*. NY: Academic Press.

Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Process, 62*(2), 198-215.

Dörner, D. (1987). On the Difficulties People Have in Dealing with Complexity. In K. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New Technology and Human Error* (pp. 97-109). Chichester: John Wiley & Sons.

Edwards, W. (1962). Dynamic decision theory and probablistic information processing. *Human Factors, 4*, 59-73.

Feiner, S., MacIntyre, B., & Seligmann, D. (1992). *Annotating the Real World with Knowledge-Based Graphics on a See-Through Head Mounted Display.* Paper presented at the Graphics Interface 1992, Vancouver, Canada.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Fong, G. T., & Nisbett, R. E. (1991). Immediate and Delayed Transfer of Training Effects in Statistical Reasoning. *Journal of Experimental Psychology, 120*(1), 34-45.

Forrester, J. W. (1968). *Principles of Systems*. Cambridge MA: Productivity Press.

Frensch, P. A., & Funke, J. (1995). Definitions, Traditions, and a General Framework for Understanding Complex Problem Solving. In P. Frensch & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 3-25). NJ: Lawrence Erlbaum Associates Publishers.

Funke, J. (1995). Experimental research on complex problem solving. In P. Frensch & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 243-268). New Jersey: Lawrence Erlbaum Associates Publishers.

Gargan, R. A., Sullivan, J. W., & Tyler, S. W. (1988). *Multimodal Response Planning: An Adaptive Rule Based Approach.* Paper presented at the CHI '88 Human Factors in Computing Systems, New York.

General Experimental Psychology Cognitive Dissonance Lab. (2002). *Cognitive Dissonance.* Available: http://www.ithaca.edu/faculty/stephens/cdback.html [2002, 22 March].

Halff, H. M. (1988). Curriculum and Instruction in Automated Tutors. In M. C. Polson & J. J. Richardson (Eds.), *Intelligent Tutoring Systems* (pp. 79-108). Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct Comparison of the Efficacy of Intuitive and Analytical Cognition in Expert Judgment. *IEEE Transactions on Systems, Man, and Cybernetics, 17*, 753-770.

Hayes, N. A., & Broadbent, D. E. (1988). Two Modes of Learning for Interactive Tasks. *Cognition, 28*, 249-276.

Hernandez, K. (1991). *Learning in Real Estate: The Role of the Development System in Creating Oversupply.* Unpublished MS Thesis, Massachusetts Institute of Technology, Boston.

Hogarth, R. M. (1981). Beyond Discrete Biases: Functional and Dysfunctional Aspects of Judgmental Heuristics. *Psychological Bulletin, 90*(2), 197-217.

Howie, E., Sy, S., Ford, L., & Vicente, K. J. (2000). Human-computer interface design can reduce misperceptions of feedback. *System Dynamics Review, 16*(3), 151-171.

Hsiao, N. (1999). *In search of theories of dynamic decision making: A literature review.* Paper presented at the International System Dynamics Conference, Wellington, New Zealand.

Internet Source. (2001). *Bloom's Taxonomy*. Available: http://www.coun.uvic.ca/learn/program/hndouts/bloom.html. Adapted from: Bloom, B.S. (1956) Taxonomy of Educational Objectives: The classification of Educational Goals: Handbook I, Cognitive Domain. New York ; Toronto: Longmans, Green. [2001, 21 September].

Jansson, A. (1995). Strategies in Dynamic Decision Making: Does Teaching Heuristic Strategies By Instructions Affect Performance. In e. a. J. P. Caverni (Ed.), *Contributions to Decision Making* (pp. 213-232). NY: Elsevier Science.

Kampmann, C. E. (1992). *Feedback Complexity and Market Adjustment: An Experimental Approach.* Unpublished Ph. D., M. I. T.

Kemp, J. E., & Dayton, D. K. (1985). *Planning and Producing Instructional Media.* New York: Harper & Row, Publishers.

Kleinmuntz, D. N. (1985). Cognitive Heuristics and Feedback in a Dynamic Decision Environment. *Management Science, 31*, 680-702.

Kleinmuntz, D. N. (1987). Human decision processes: Heuristics and task structure. In P. A. Hancock (Ed.), *Human Factors Psychology* (pp. 123-157). New York: Elsevier Science.

Kleinmuntz, D. N. (1993). Information Processing and Misperception of the Implications of Feedback in Dynamic Decision Making. *System Dynamics Review, 9*(3), 223-237.

Kleinmuntz, D. N., & Kleinmuntz, B. (1981). Systems Simulation: Decision Strategies in Simulated Environments. *Behavioral Science, 26*, 294-305.

Kluwe, R. H. (1995). Single Case Studies and Models of Complex Problem Solving. In P. Frensch & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 269-291). NJ: Lawrence Erlbaum Associates Publishers.

Kondratiev, N. (1935). The long waves of economic life. *Review of Economic Statistics, 17*, 105-115.

Locke, E. A., & Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance.* New York: Prentice-Hall.

Mackinnon, A. J., & Wearing, A. J. (1980). Complexity and Decision Making. *Behavioral Science, 25*(4), 285-296.

Marks, J. W. (1991). Discourse Coherence and the Consistent Design of Informational Graphics. In M. T. Maybury (Ed.), *Working Notes from the AAAI Workshop on Intelligent Multimedia Interfaces* (pp. 29-36). Menlo Park, California: AAAI.

Maxwell, T. A. (1995). *Decisions: Cognitive Styles, Mental Models, and Task Performance.* Unpublished Ph.D. Dissertation, State University of New York at Albany, Albany.

McGeorge, P., & Burton, A. M. (1989). The Effects of Concurrent Verbalization on Performance in a Dynamic Systems Task. *British Journal of Psychology, 80*, 455-465.

Paich, M., & Sterman, J. D. (1993). Boom, Bust, and Failures to Learn in Experimental Markets. *Management Science, 39*(12), 1439-1458.

Price, R. V. (1991). *Computer-Aided Instruction: A Guide for Authors*. Pacific Grove, California: Brooks/Cole Publishing Company.

Richardson, G. P. (1991). *Feedback Thought in Social Science and Systems Theory*. Philadelphia: University of Pennsylvania Press.

Richardson, G. P., & Rohrbaugh, J. (1990). Decision Making in Dynamic Environments: Exploring Judgments in a System Dynamics Model-Based Game. In K. Borcherding, O. I. Larichev, & D. M. Messick (Eds.), *Contemporary Issues in Decision Making* (pp. 463-472). Amsterdam: North-Holland.

Roth, S. F., & Hefley, W. E. (1993). Intelligent Multimedia Presentation Systems: Research and Principles. In M. T. Maybury (Ed.), *Intelligent Multimedia Interfaces* (pp. 13-58). Menlo Park, California, Cambridge, Massachusetts: American Association for Artificial Intelligence Press, Massachusetts Institute of Technology Press.

Roth, S. F., & Hendrickson, C. T. (1991). Computer Generated Explanations in Project Management Systems. *Journal of Computing in Civil Engineering, 5*(2), 231-244.

Roth, S. F., & Mattis, J. (1990). *Automatic Graphics Presentation for Production and Operations Management Systems.* Paper presented at the Fourth International Conference on Expert Systems for Production and Operations Management, Hilton Head, South Carolina.

Sanderson, P. M. (1989). Verbalizable Knowledge and Skilled Task Performance: Association, Dissociation, and Mental Model. *Journal of Experimental Psychology: Learning Memory and Cognition, 15*, 729-747.

Schwier, R. A., & Misanchuk, E. R. (1993). *Interactive Media Instruction*. Englewood Cliffs, New Jersey: Educational Technology Publications.

Sengupta, K., & Abdel-Hamid, T. K. (1993). Alternative Conceptions of Feedback in Dynamic Decision Environments: An Experimental Investigation. *Management Science, 39*(4), 411-428.

Shanteau, J. (1992). Competence in Experts: The Role of Task Characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252-266.

Soulier, J. S. (1988). *The Design and Development of Computer Based Instruction*. Boston: Allyn and Bacon, Inc.

Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (1989). Insight without Awareness: On the Interaction of Verbalization, Instruction and Practice in a Simulated Process Control Task. *The Quarterly Journal of Experimental Psychology, 41A*(3), 553-577.

Steinberg, E. R. (1991). *Computer-Assisted Instruction.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Sterman, J. D. (1987). Testing behavioral simulation models by direct experiment. *Management Science, 33*, 1572-1592.

Sterman, J. D. (1989a). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes, 43*(3), 301-335.

Sterman, J. D. (1989b). Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Management Science, 35*(3), 321-339.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review, 10*(2-3), 291-330.

Sterman, J. D., & Meadows, D. (1985). STRATEGEM-2: A microcomputer simulation game of the Kondratiev Cycle. *Simulation and Games, 16*(2), 174-202.

Sternberg, R. J. (1995). Expertise in Complex Problem Solving: A Comparison of Alternative Conceptions. In P. Frensch & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 295-321). NJ: Lawrence Erlbaum Associates Publishers.

Stewart, T. R., & Hsiao, N. (1997). *Judgmental accuracy and task predictability.* Paper presented at the Thirteenth Annual International Invitational Meeting of the Brunswik Society.

Trees, S., Doyle, J., & Radzicki, M. (1996). *Using Cognitive Styles Typology to Explain Dynamic Decision Making in a Computer Simulation Game Environment.* Paper presented at the *International System Dynamics Conference.*

Vicente, K. J. (1996). Improving Dynamic Decision Making in Complex Systems through Ecological Interface Design: A Research Overview. *System Dynamics Review, 12*(4), 251-279.

Wang, S. (1994). *Learning Laboratory Design and Learning Transfer.* Unpublished Ph.D. Dissertation (in Chinese), National Sun Year-End University, Causing, Taiwan.

Yang, J. (1996). *Facilitating Learning through Goal Setting in A Learning Laboratory.* Paper presented at the *International System Dynamics Conference.*

Yang, J. (1997). *Give Me the Right Goals, I Will Be A Good Dynamic Decision Maker.* Paper presented at the International System Dynamics Conference.

Young, S. H., Chen, C. P., Wang, S., & Chen, C. H. (1997). *An Experiment to Study the Relationship between Decision Scope and Uncontrollable Positive Feedback Loops.* Paper presented at the International System Dynamics Conference.

APPENDIX A

Dependent Variables for Dynamic Decision Making[†]

Category	Measure and Studies
Task Performance	**Optimizing, maximizing or minimizing, specified measures or benchmarks** – *Cost, the higher the cost, the lower the performance* (Sterman, 1989a; Sterman, 1989b; Richardson & Rohrbaugh, 1990; Wang, 1994; Diehl et al., 1995; Maxwell, 1995; Trees et al., 1996) – *Profit* (Kampmann, 1992; Bakken, 1993; Paich et al., 1993; Yang, 1996; Young et al., 1997) – *Patients' health conditions* (Kleinmuntz et al., 1981) – *Proportion of patients cured* (Kleinmuntz, 1985; Kleinmuntz et al., 1987) – *Number (percent) of areas lost* (Brehmer, 1990; Brehmer et al., 1991; Brehmer, 1995; Brehmer et al., 1995) – *Difference* (percent difference) compared with a benchmark (Broadbent et al., 1978; Mackinnon et al., 1980; Bakken, 1993) – *Number of decision outcomes better than a benchmark* (Broadbent et al., 1986) **Reaching specified targets** – *Number (percent) of attempts within a range of a specified target* (Berry et al., 1984; Broadbent et al., 1986; Berry et al., 1987; Berry et al., 1988; Hayes et al., 1988; McGeorge et al., 1989; Sanderson, 1989; Stanley, et al., 1989) – *Number (percent) of attempts in correct directions to reach the target* (Sanderson, 1989; Yang, 1996) – *Number (percent) of errors of directions to reach the target* (Broadbent et al., 1986; Berry et al., 1987) **Task systems behaviors** – *Number of systems destruction* (Yang, 1997) – *Number of appearances of an archetype "fixes that fail"* (Yang, 1997) **Goals combining two criteria** – *Market share and cumulative net marketing contribution (consistent goals)* (Hogarth et al., 1981) – *Cost and schedule (conflicting goals)* (Sengupta et al., 1993) **Goals combining multiple (greater than two) criteria** – *A composite index based on six indicators* (Jansson, 1995)

[†] NOTE: The table in this appendix has been adapted/modified from {Hsiao, N. (1999). *In search of theories of dynamic decision making: A literature review.* Paper presented at the International System Dynamics Conference, Wellington, New Zealand.}.

APPENDIX A – Con't

Learning	**Mean scores of pre-game and/or post-game questionnaires on the relationships of variables, including those direct and crossed relationships between variables** – *Declarative task knowledge* (Broadbent et al., 1978; Berry et al., 1984; Broadbent et al., 1986; Berry et al., 1987; Berry et al., 1988; Hayes et al., 1988; Sanderson, 1989; Bakken, 1993; Jansson, 1995; Maxwell, 1995; Trees et al., 1996; Howie et al., 2000) **Mean scores of pre-game and/or post-game questionnaires same as above** – *Procedural task knowledge* (Hayes, et al., 1988) **Number of correctness of mental models aligned with heuristics and goals set forth** – (Yang, 1996; Yang, 1997) **Number matching certain types mental models** – (Sanderson, 1989) **Performance on transferred tasks** – (Berry et al., 1988; Hayes et al., 1988; Bakken, 1993; Wang, 1994)
Efforts for Decision Making	**Amounts of decision time** – (Kleinmuntz et al., 1987; Sanderson, 1989; Brehmer, 1990; Brehmer et al., 1991; Sengupta et al., 1993; Wang, 1994; Brehmer, 1995; Brehmer et al., 1995; Diehl et al., 1995; Jansson, 1995; Maxwell, 1995; Yang, 1996) **Amounts of information use for specific information items** – (Brehmer et al., 1991; Sengupta et al., 1993; Brehmer et al., 1995; Jansson, 1995; Maxwell, 1995; Yang, 1996) **Amounts of discussion among subjects** – (Hogarth et al., 1981)
Quality of Decision-Making Process	**Decision scope** – *Number of different decision rules employed* (Wang, 1994; Young et al., 1997) **Reliability** – *Fluctuations of decisions* (Sengupta et al., 1993)
Decision-Making Architecture	**Delegation of decision making** – (Brehmer et al., 1991)

APPENDIX B

Independent Variables for Dynamic Decision Making[†]

Category	Conceptual Definition	Measures and Studies
Decision-Maker Factors	**Cognitive style (ability)**	- *MBTI (Myers-Briggs Type Indicator)* (Maxwell, 1995; Trees et al., 1996) - *Gregoric Style Delineator (four mediation channels)* (Trees et al., 1996) - *Gordon's Cognitive Style Indicator (four types)* (Trees et al., 1996)
	Task Expertise/ Academic Training	- *Whether subjects have task domain expertise in terms of their academic background* (Bakken, 1993) - *Whether subjects receive a 2-day session involving simulation of the JOBS program* (Maxwell, 1995)
	Computing skills	- *Subjects' self-rating evaluation about their computer use skills* (Trees et al., 1996)
	Practice / task experience	- *Whether subjects experience repeated trials (not explicitly manipulated)* (Broadbent et al., 1978; Kleinmuntz et al., 1987; Berry et al., 1987; Berry et al., 1988; Stanley et al., 1989; Brehmer, 1990; Brehmer et al., 1991; Bakken, 1993; Sengupta et al., 1993; Paich et al., 1993; Wang, 1994; Diehl et al., 1995) - *Amounts of practice from repeated trials* (Berry et al., 1984; Broadbent et al., 1986; Sanderson, 1989) - *Whether subjects experience a conceptually similar task for the next trial block* (Berry et al., 1988)
Task Complexity	**Total variables**	- *Total number of variables in task systems* (Mackinnon et al., 1980)
	Interaction b/w subsystems	- *Whether interaction exists between variables or subsystems* (Mackinnon et al., 1980)
	Random variation	- *Whether random variation exists at strategic points in tasks* (Mackinnon et al., 1980)
	Miscellaneous task characteristics	- *Initial health, treatment risk, and symptom diagnosticity* (Kleinmuntz, 1985) - *Treatment risk (Appearance or strength)* (Kleinmuntz et al., 1987) - *Levels of price regime* (Kampmann, 1992) - *Types of software project* (Sengupta et al., 1993)
	Time delay / lagged effects	- *Lagged effects* (Broadbent et al., 1978; Broadbent et al., 1986; Berry et al., 1988; Paich et al., 1993) - *Time constants* (Sterman, 1989a; Sterman, 1989b; Brehmer, 1990; Brehmer et al., 1991; Kampmann, 1992; Brehmer, 1995; Diehl et al., 1995)

[†] See NOTE to Appendix A. Same source and modifications apply here.

APPENDIX B Con't

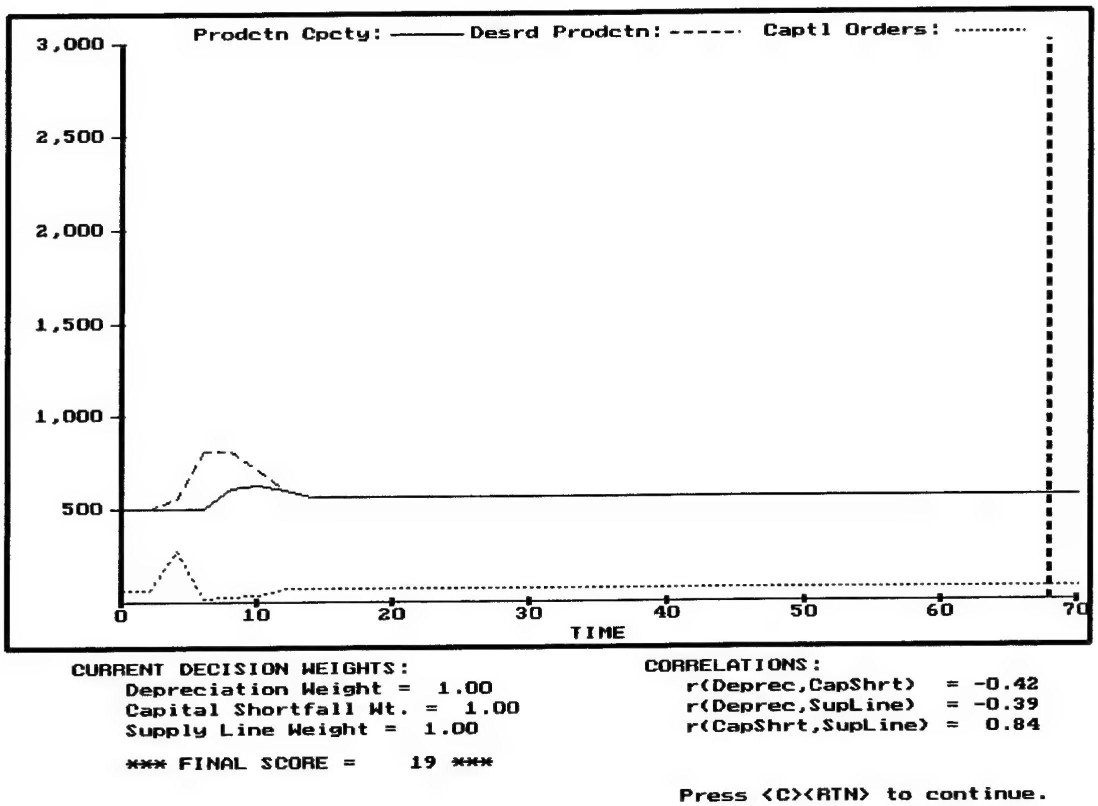
	Effectiveness of decisions on outcomes	- *Treatment effectiveness* (Kleinmuntz, 1985) - *Reducing stability by enlarging effects of a decision on outcomes* (Broadbent et al., 1986) - *Effectiveness of fire-fighting units* (Brehmer et al., 1991)
	Frequency of oscillation	- *Number of peaks of prices* (Bakken, 1993)
	Positive feedback and gains (appearance or strength)	- *Positive gains built in the task model* (Sterman, 1989a; Sterman, 1989b; Kampmann, 1992; Diehl et al., 1995) - *Strength of "word of mouth"* (Paich et al., 1993) - *Number of intervals a system falls in the uncontrollable positive loops* (Young et al., 1997)
	Real-time tasks	- *Whether a task system is clock-driven or event-driven* (Brehmer, 1995)
Decision-Making Interfaces and Environments	**Heuristics (decision rules) built in task systems**	- *3 levels: 1) arbitrary consistent, 2) arbitrary-random, and 3) none (left for human judgment)* (Hogarth et al., 1981) - *3 levels of strategies with increasing computational complexity: 1) generate-and-test, 2) heuristic, and 3) EU-bayesian* (Kleinmuntz et al., 1981) - *Random vs. schema-driven strategies, 2 levels of information acquisition, 2 levels of base-rate utilization, 3 levels of computational complexity* (Kleinmuntz, 1985)
	Learning via lagged effects	- *Selective-mode or unselective mode by varying lagged effects of decisions* (Hayes, et al., 1988)
	Heuristics induced goal setting that subjects receive through verbal directions	- *2 types: 1) total assets goal (long-term whole-system goal) and 2) total assets and order growth goal (short-term subsystem goal)* (Yang, 1996) - *3 types: prey/predator (whole-system) ratio, prey/predator (whole-system) number, and prey (sub-system) number* (Yang, 1997)
	Task property, strategies, and heuristics (decision rules) that subjects receive through verbal instructions	- *Training / no training concerning task property* (Berry et al., 1984; Berry et al., 1987; Berry et al., 1988) - *3 levels of task property: 1) no preliminary training, 2) trained with relationships of variables, and 3) practicing each pair of relationships separately* (Broadbent et al., 1986) - *3 levels of expert transcripts: 1) no transcript, 2) block-by-block transcript, and 3) whole transcript* (Stanley et al., 1989) - *5 levels of instructions: 1) no training, 2) expert transcript, 3) memory training, 4) rule construction, 5) simple rule* (Stanley et al., 1989) - *5 types of expert transcripts: 1) no training, 2) initial blocks, 3) final blocks, 4) pre-cutpoint of performance, 5) post-cutpoint of performance* (Stanley et al., 1989) - *2 levels of instructions: 1) systematic-elaborate: variables' relationship, 2) goal-planning: detailed measures of decisions and outcomes* (Jansson, 1995) - *3 levels of training: 1) causal loop, 2) strategic time plots, and 3) strategic heuristics)* (Maxwell, 1995)

APPENDIX B Con't

	Concurrent verbalization / thinking-aloud	- *Whether, while playing the game, subjects are required to verbally describe tasks and heuristics employed* (Berry et al., 1984; McGeorge et al., 1989; Stanley et al., 1989)
	Increasing task salience	- *Between trial blocks, instruct subjects with task structures and effects of decisions and time delay* (Berry et al., 1988; Wang, 1994) - *Whether subjects are informed with appearance of delay* (Brehmer, 1995)
	Degree of dec'n precision req'd	- *Whether subjects are required to place decisions to the first decimal place* (Sanderson, 1989)
	Learning inducement	- *Prior to tasks, instruct subjects to focus on searching for task pattern and structure* (Berry et al., 1988) - *Prior to tasks, induce learning by instructing subjects that learning is crucial and task performance does not affect economic reward* (Wang, 1994)
	Contents of information display	- *Whether Bayesian strategy is available* (Kleinmuntz et al., 1987) - *Whether subjects' previous decisions and outcomes are available* (Sanderson, 1989) - *3 levels: 1) Feedforward: whether the subjects learned the three formula; 2) cognitive feedback: whether the subjects received task information; 3) outcome feedback: project status reports in numerical forms* (Sengupta et al., 1993)
	Forms of information display	- *Whether subjects receive graphical representations of system status* (McGeorge et al., 1989; Sanderson, 1989) - *Whether subjects receive formula for decisions* (Sanderson, 1989) - *Whether subjects only receive variables' names without semantic meanings* (Sanderson, 1989) - *3 levels: 1) no cue highlighted, 2) all cues highlighted (cue discovery), and 3) all cues highlighted plus heuristics (feedforward)* (Richardson & Rohrbaugh, 1990)- *Improved graphics display* (Richardson & Rohrbaugh, 1990, Howie et al., 2000)
	Decision-making architectures	- *Whether subjects use hierarchical or networked decision-making* (Brehmer, et al., 1995)

APPENDIX C

Sterman "Optimal" Solution in Computer Game



Graphic taken from STRATEGEM-2 for DOS ©1985 by John Sterman

In year four, when a step increase in goods orders rises from 450 units to 500, the optimal solution produces an order of 260 units for the capital sector. In the following year, zero units are ordered. For year eight, 10 units are ordered. Year 10 orders are 20, and in year 12, 60 capital sector orders are made and remain that way for the remainder of the game – producing the optimal score of 19.

APPENDIX D

Original Experiment Instructions

Welcome to the STRATEGEM-2 Simulation Game[†]

Version 2.1 Copyright 1985 John Sterman

The economic malaise of the 1980's has revived interest in the economic long wave or Kondratiev Cycle, a cycle of prosperity and depression averaging 50 years.

Since 1975 the System Dynamics National Model has provided an increasingly rich theory of the long wave. The theory emerging from the National Model explains the long wave as the endogenous result of decision making by individuals, corporations, and government. However, the complexity of the National Model makes it difficult to explain the dynamics underlying the long wave. This game demonstrates how long waves can arise by focusing on the role of capital investment.

There are two basic kinds of industries in modern economies: capital producers and producers of consumer goods and services. Goods producers sell primarily to the public. Producers of capital make and sell the plant and equipment that the consumer sector needs in order to produce goods and services. But, in addition, the capital-producing industries of the economy (construction, heavy equipment, steel, mining, and other basic industries) supply each other with the capital, plant, equipment, and materials each need to operate. Viewed as a whole, the capital sector of the economy orders and acquires capital from itself.

You will manage the capital producing sector of the economy. Your goal is to balance the supply and demand for the capital. To do this you must keep your production capacity (current capacity) as closely matched to the demand (total backlogs) for capital as possible. The game is won by the person with the lowest score. The score is the average absolute deviation between production capacity and desired production. For example, if capacity were 500 and demand were 600, your score for that period would be 100. Likewise, if capacity were 600 and demand were only 500, your score for that period would also be 100. A Score of zero means supply and demand are in perfect balance. You are therefore penalized for excess capacity (which implies some of your factories are idle) and also for insufficient capacity (which means you are unable to meet the demand for capital).

Time is divided into two-year periods. At the beginning of each period, orders for capital are received from two sources: the goods sector and the capital sector itself.

[†] These instructions were taken from the Howie STRATEGEM-2 interface (2000).

Orders for capital arriving from the goods sectors are determined by the computer. Orders for capital you placed in the previous period are moved into the unfilled order backlog for the capital sector.

Orders placed by the goods and capital sectors accumulate in the backlog of unfilled orders for each sector. The total backlog of orders is the desired production for the current two-year period, the demand you must meet.

Production itself is the lesser of desired production or production capacity. Production capacity is determined by the capital stock of the sector. Capital stock is decreased by depreciation and increased by shipments. You lose 10% of your stock each period.

If capacity is inadequate to meet demand fully, available production of capital is allocated between the capital and goods sectors in proportion of their respective backlogs. For example, if the backlog from the capital sector were 500 and the backlog from the goods sector were 1000, desired production would be 1500.

If capacity were only 1200, production would be 1200 and the fraction of demand satisfied would be 1200/1500 = 80%. Thus 400 units would be shipped to the capital sector and 800 would be shipped to the goods sector.

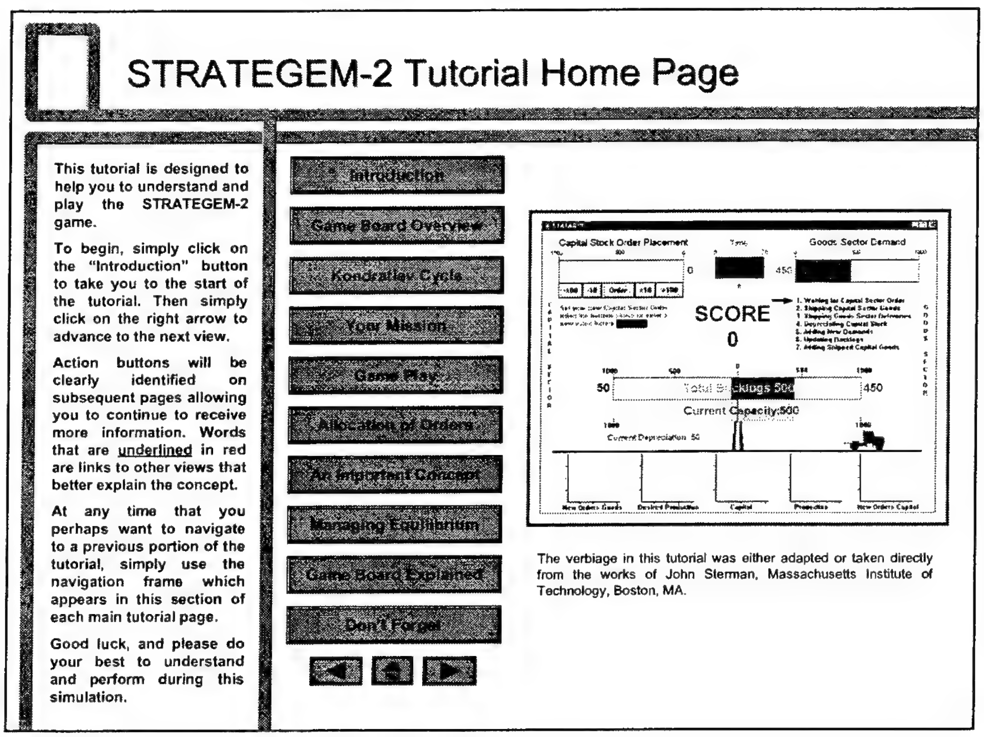
Any unfilled orders remain in their respective backlogs to be filled in future periods. In the example, 100 units would remain in the backlog of the capital sector and 200 would remain in the backlog of the goods sector.

APPENDIX E

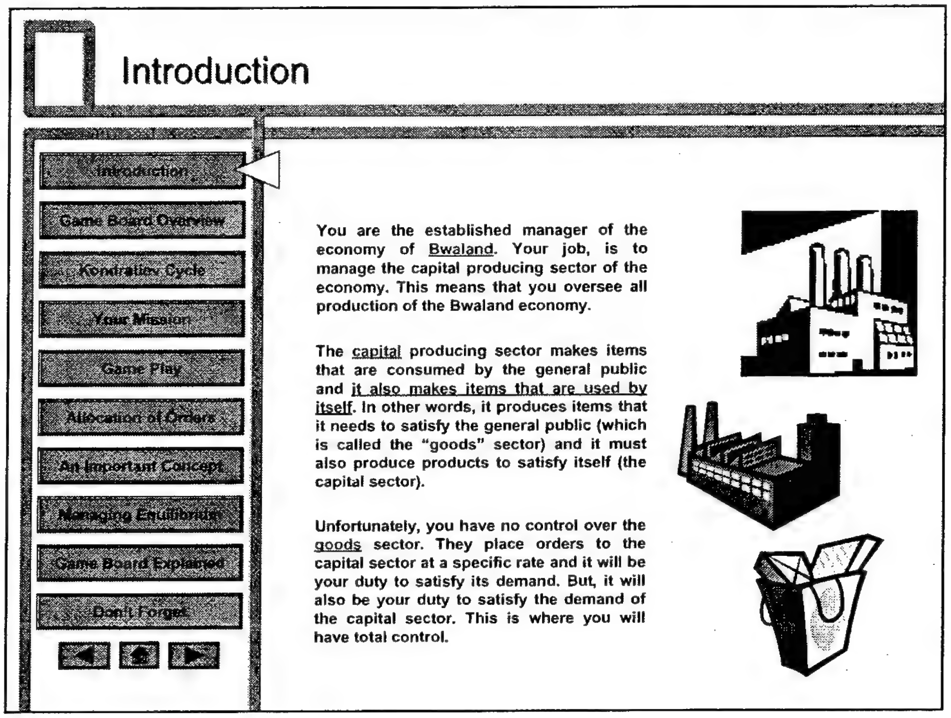
Bois Instructions

This section will be an on-screen tutorial provided to participants. Below are pertinent views of the tutorial.

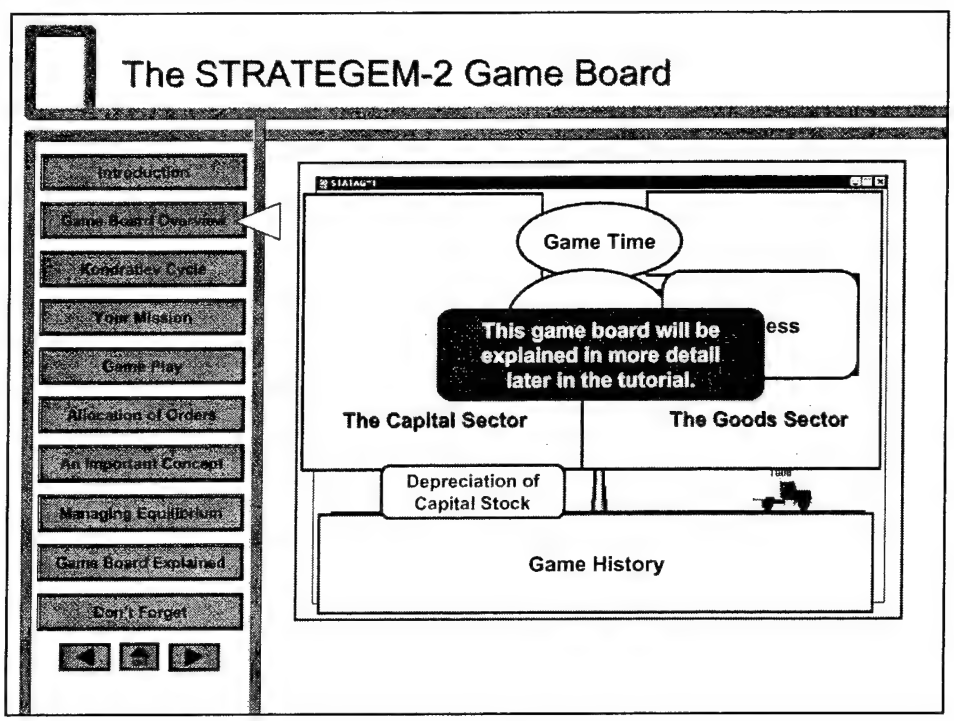
First page: Contains navigation instructions.



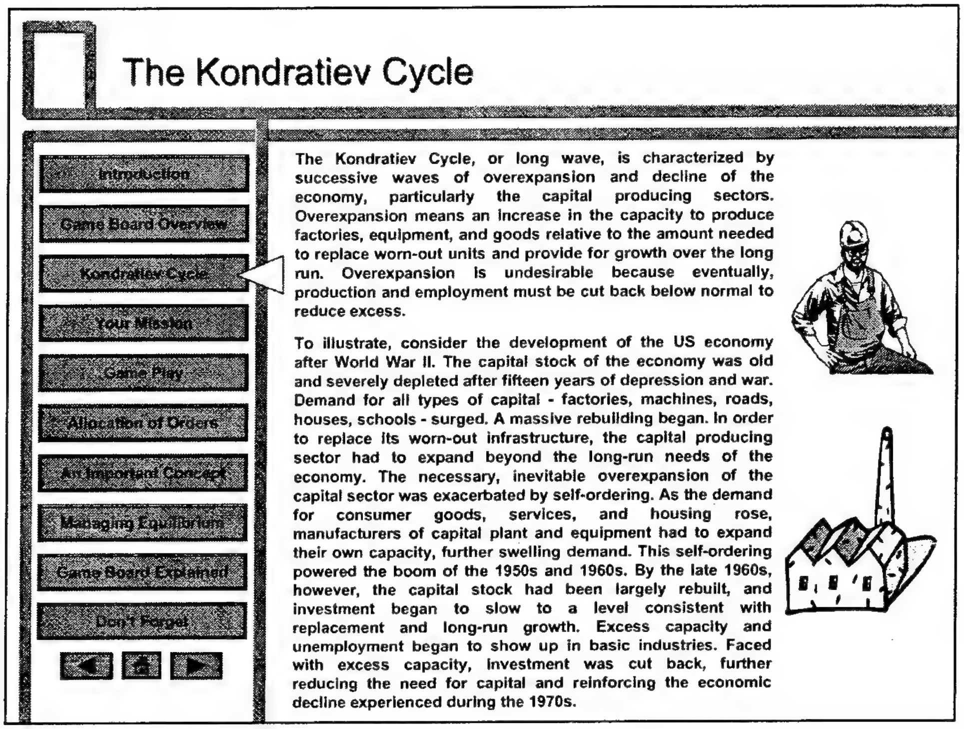
Introduction View: Contains links to other hidden pages to further define terms.



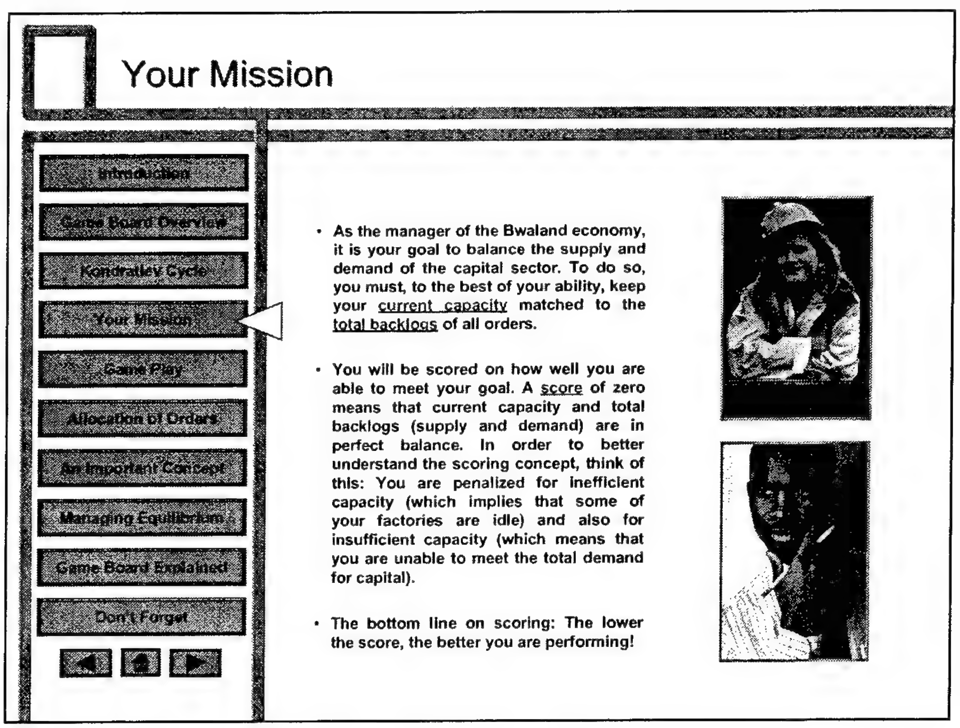
Game board overview: Has multiple overlays used to familiarize participant with game board



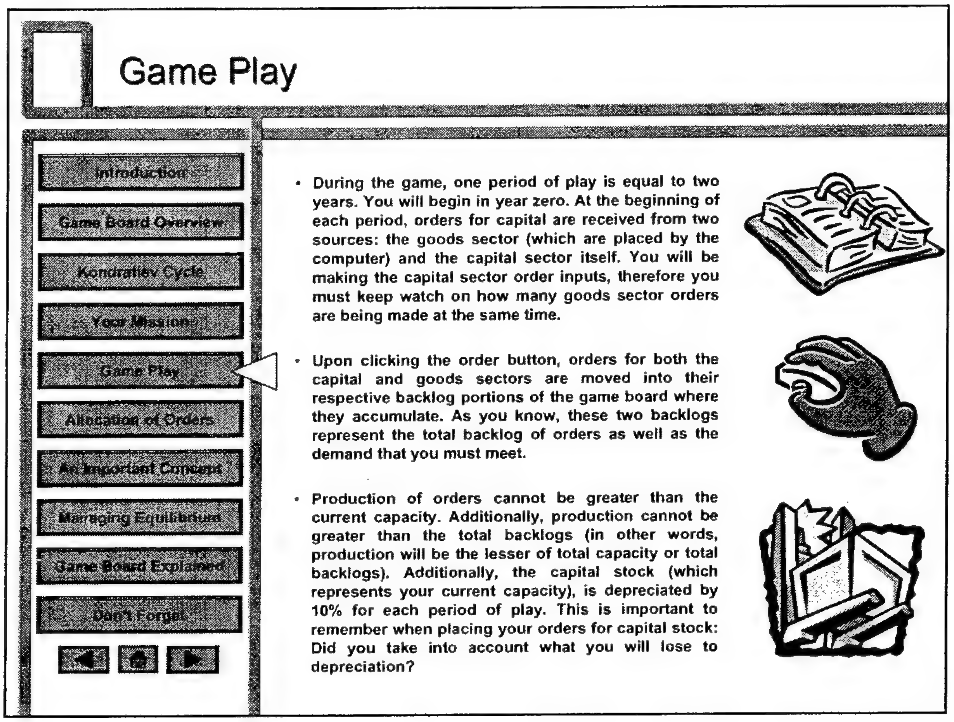
Fourth view: Explains the Kondratiev Cycle



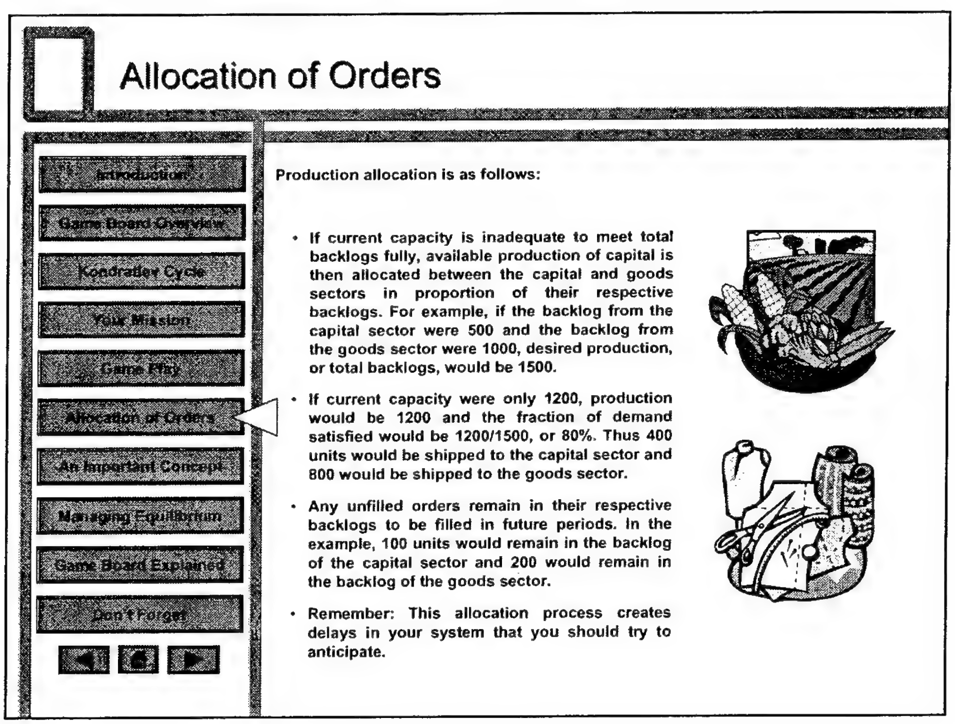
Fifth view: Explains goal / scoring of simulation with links to other explanations.



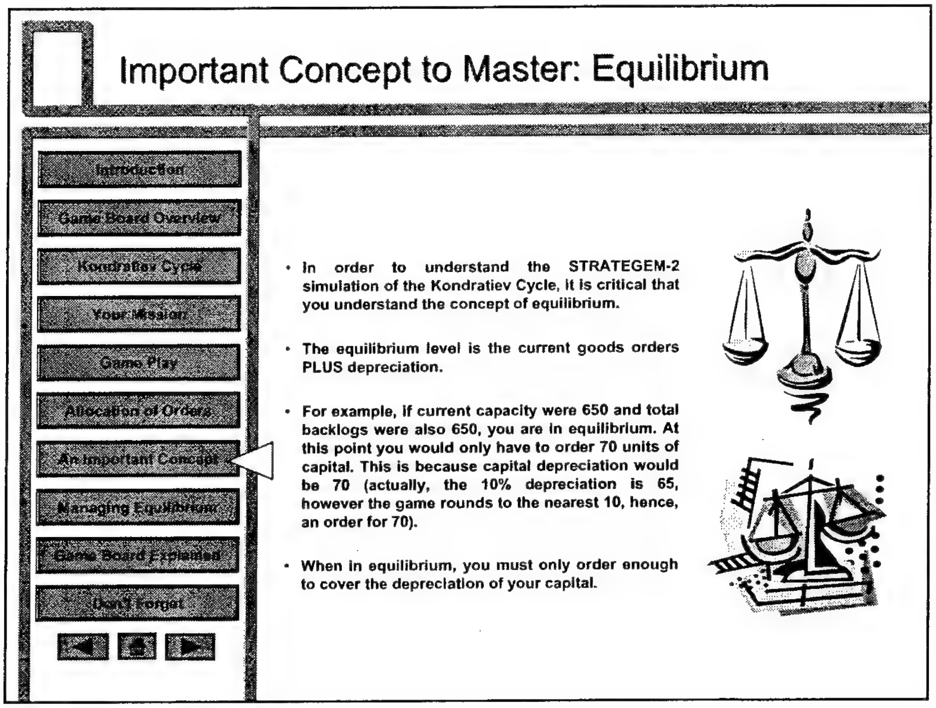
Sixth view: Outlines the play of the game.



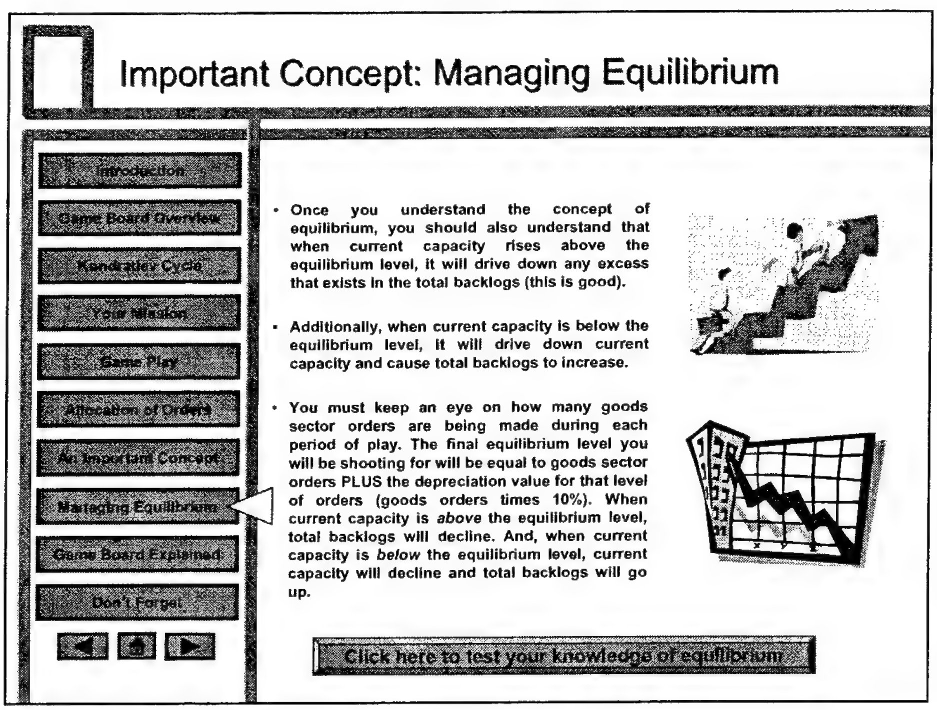
Seventh view: Explains how order allocations are made to each sector.



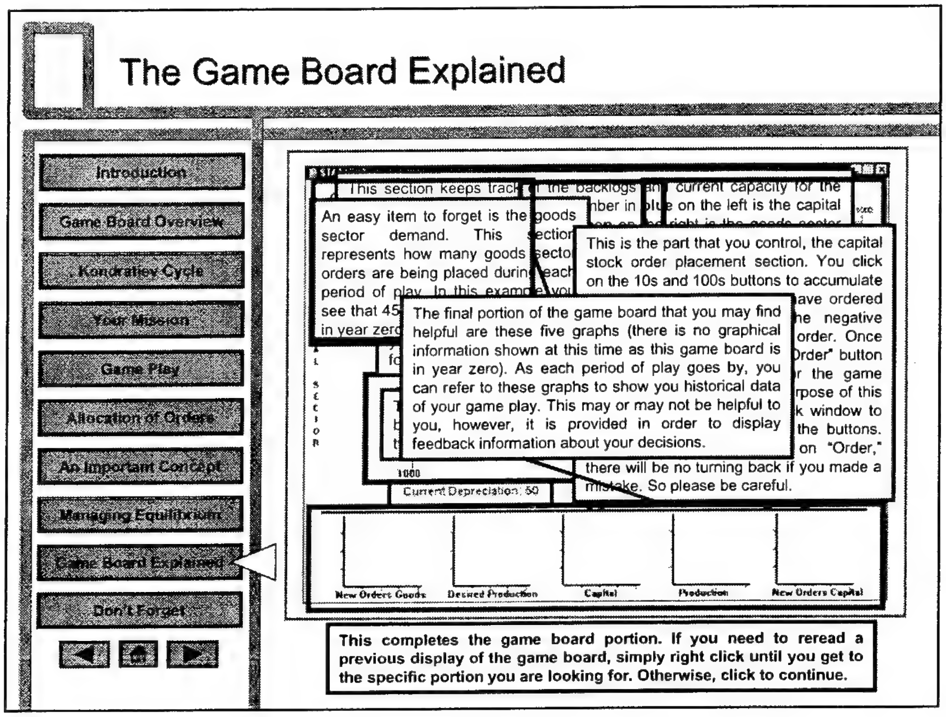
Eighth view: Introduces the concept behind game equilibrium.



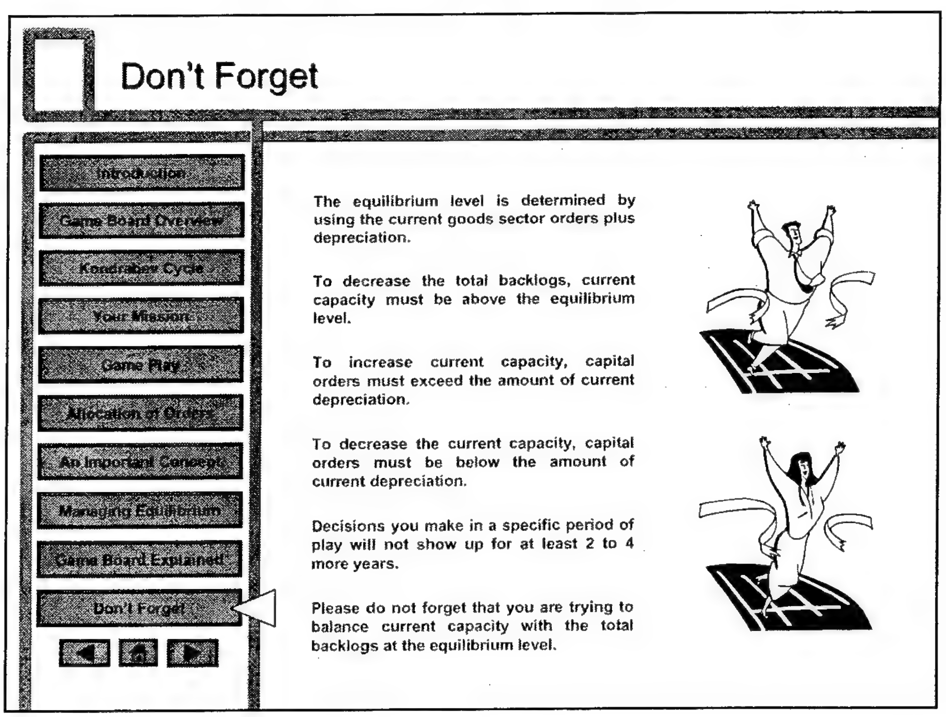
Ninth view: Explains how one manages equilibrium.
Includes link to 4-question exam used to bolster learning (not shown)



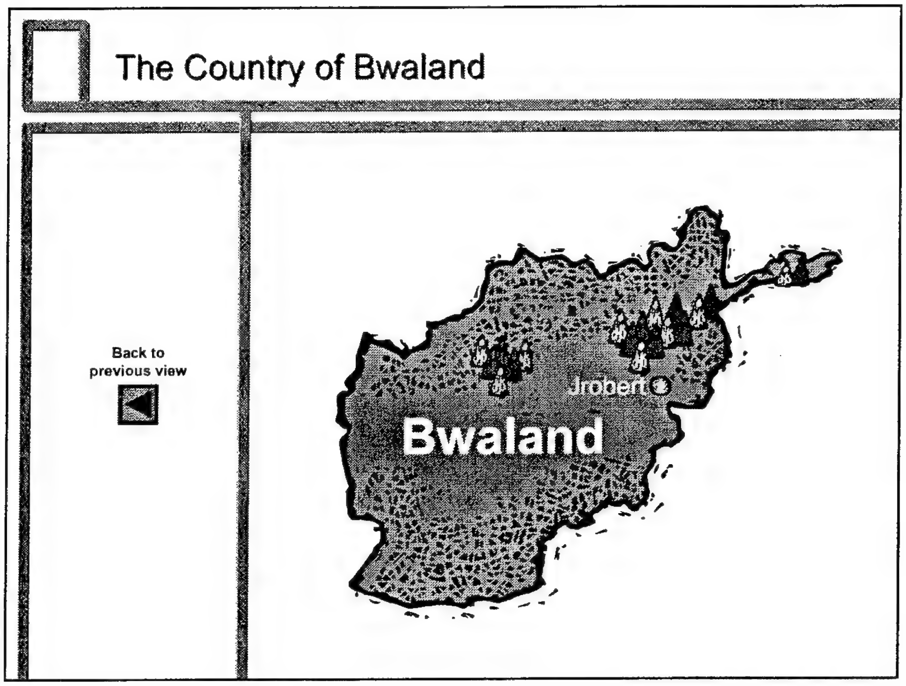
Tenth view: Explains the sections of the game board – a multiple view display.



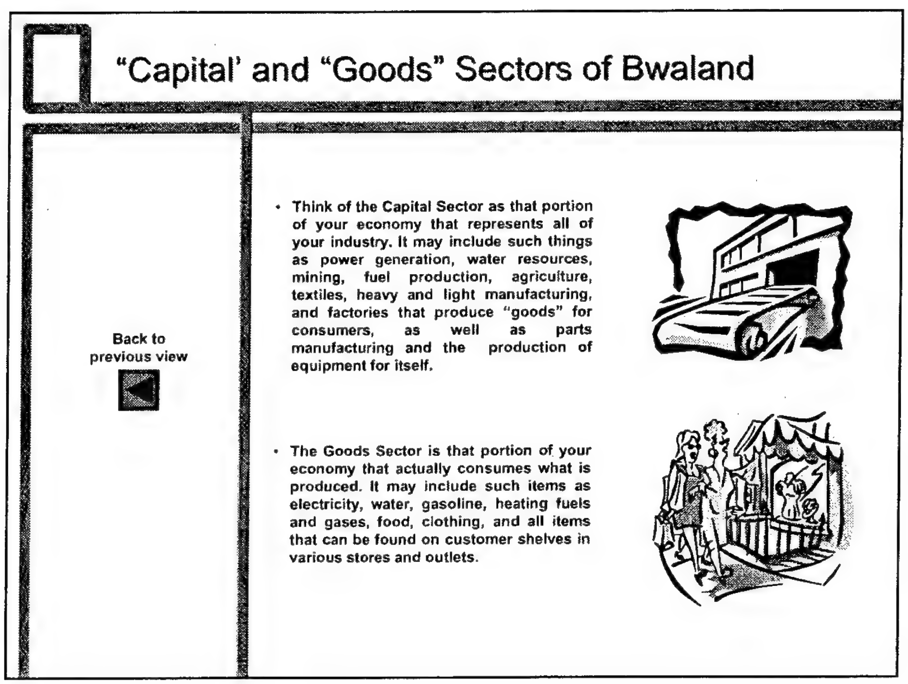
Final view: Provides tips to remember.



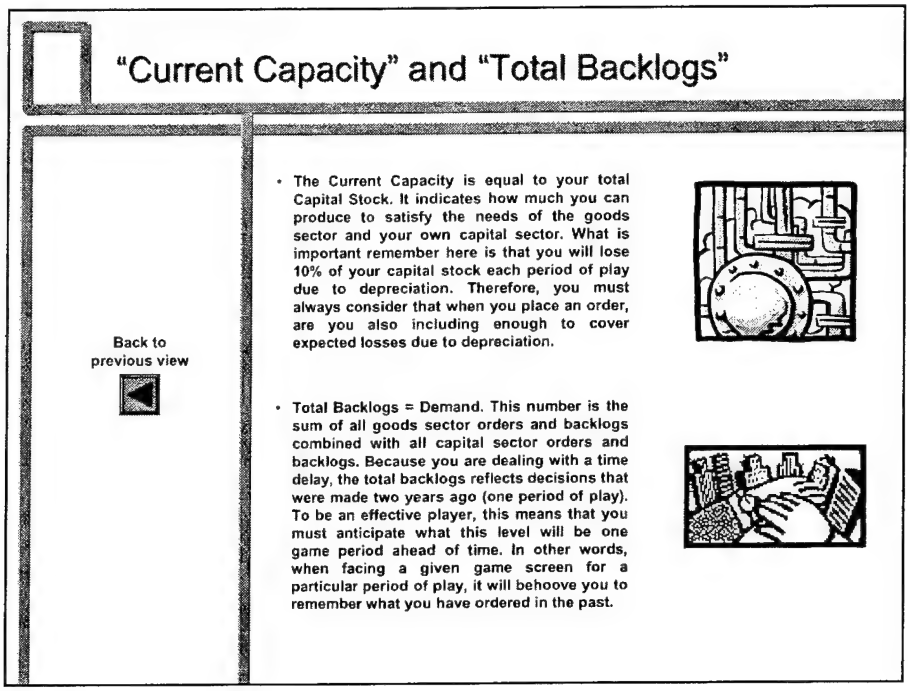
First linked view: Accessed only from another page. Used to define Bwaland.



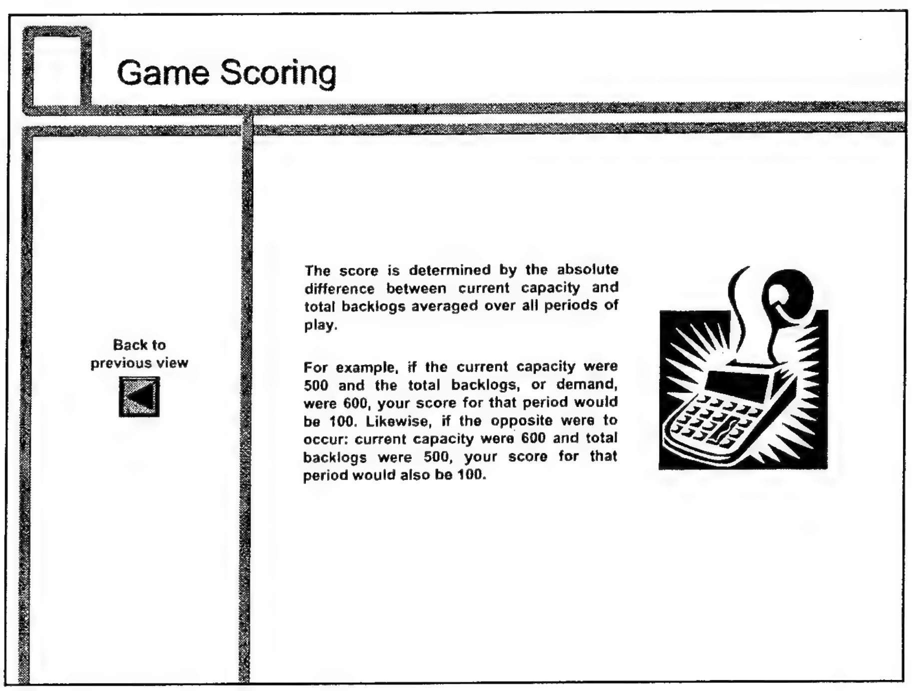
Second linked view: Used to further define the capital and goods sectors.



Third linked view: Used to further define current capacity and total backlogs.



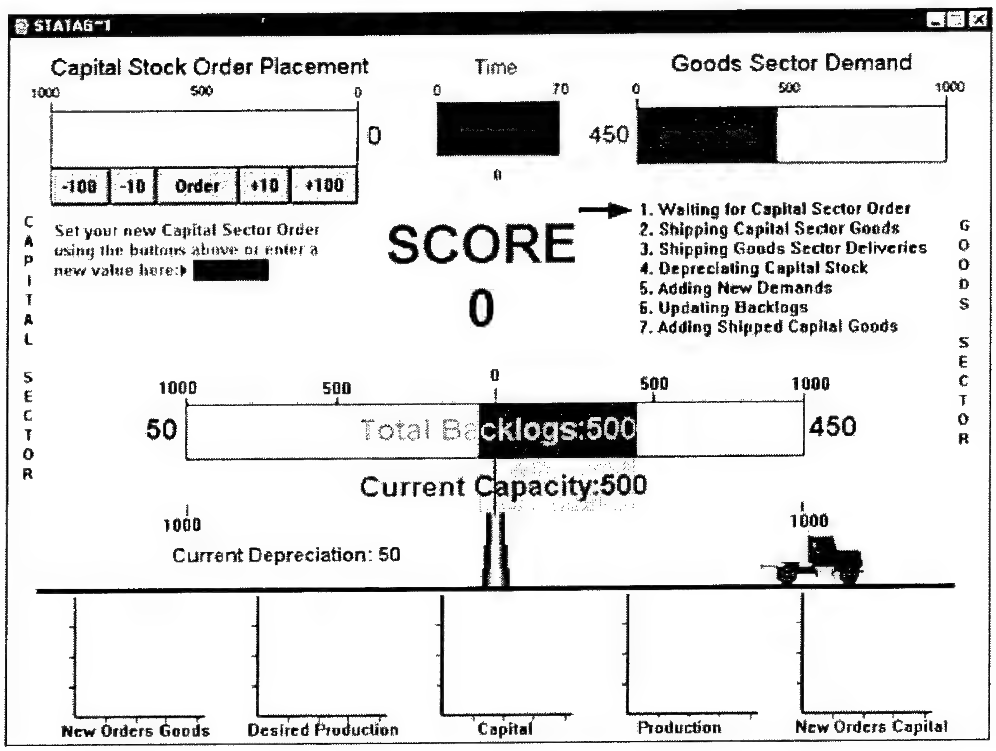
Fourth linked view: Explains how the game is scored.



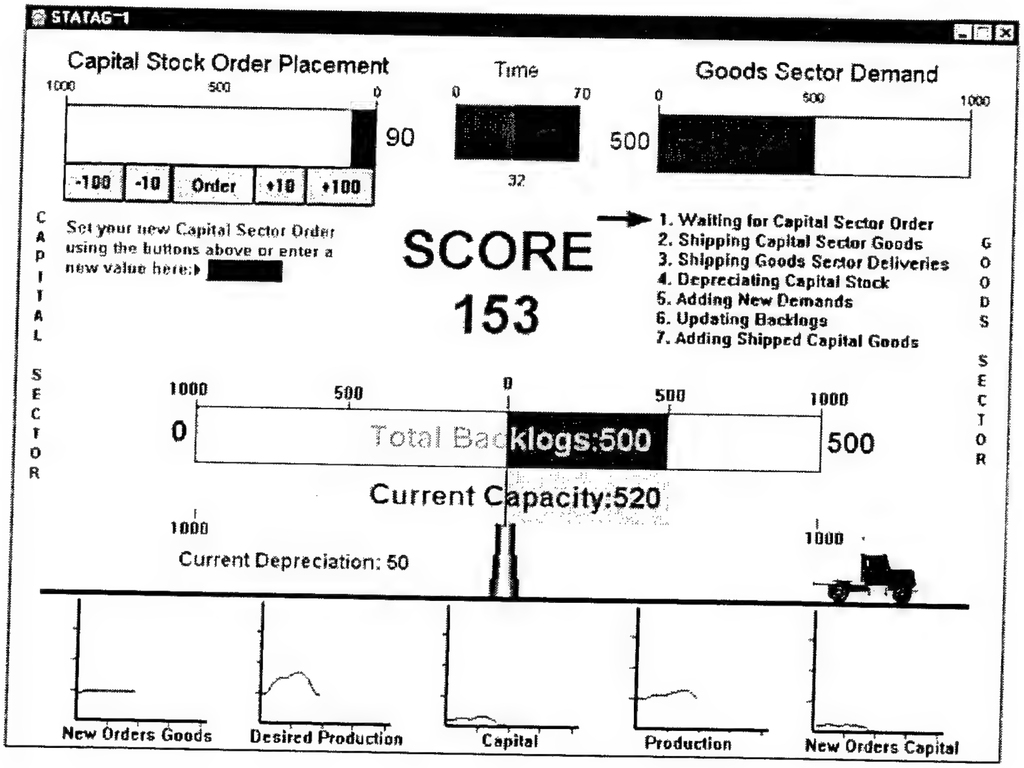
APPENDIX F

The Howie STRATEGEM-2 Interface

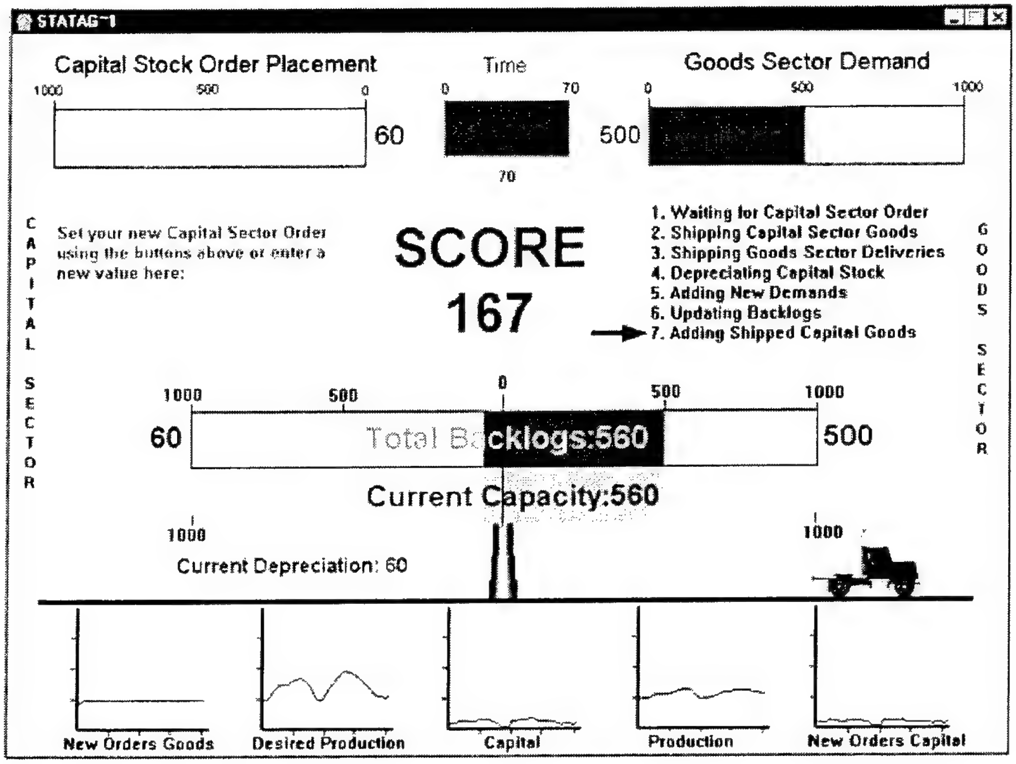
The following depiction of the STRATEGEM-2 interface shows the beginning of the game. It indicates a goods sector demand of 450 orders with an overall capacity of 500. The participant would need to only order 50 units in the capital sector (which is just enough to accommodate depreciation). The 50 capital orders combined with the 450 goods sector orders equals the 500 units of total capacity and would therefore keep the game in equilibrium and keep the score at zero.



The depiction shown below is in year 32 of a sample game played by the researcher. At this point, the researcher has allowed the current capacity (520) to be too low in order to meet the total demand, or desired production (500) plus accommodate depreciation (50). Therefore, current capacity should at least be 550 to maintain the game in equilibrium at this point. The depiction below occurred from under-ordering in the previous timeframe. In response, the researcher is ordering 90 capital units in order to boost capacity in future years.



This final depection shows the sample game in the final year (year 70). The researcher has managed to get the game back in equilibrium (this occurred in year 68). At this point, orders required for the capital sector need only to accommodate the current depreciation (60 units). The final score for the game is 167.



Knowledge Survey

Adapted from: Howie, E., Sy, S., Ford, L., & Vicente, K. J. (2000). Human-computer interface design can reduce misperceptions of feedback. *System Dynamics Review, 16*(3), 151-171.

Participant Number _____________

1. Is there depreciation on "goods" shipped to the goods sector?

   a) Yes

   b) No

2. What is(are) the main sector(s) in the economy?

   a) Depreciation

   b) Goods

   c) Capital

   d) A and C only

   e) B and C only

   f) A and B only

3. Of which sectors (use question 2's options) do you have control over?

   a) Depreciation

   b) Goods

   c) Capital

   d) A and C only

   e) B and C only

   f) A and B only

4. If the current capacity increases and any other demands stay the same, the subsequent total backlog:

   a) Increases

   b) Decreases

   c) Does not change

   d) The current capacity is irrelevant to the backlog

5. The Kondratiev long wave is used to depict the results of over-expansion in a capital sector:

   a) True

   b) False

6. Total Backlog consists of:

   a) Capital sector orders

   b) Capital depreciation

   c) Goods sector orders

   d) All of the above

   e) A and C only

   f) None of the above

7. The factor(s) that contributes to the game score is(are):

   a) Current Capacity

   b) Total Backlogs

   c) The Game Period

   d) All of the above

   e) A and B only

8. To get the best score, you should:

   a) Maximize overproduction

   b) Minimize over and underproduction

   c) Maximize underproduction

   d) None of the above

9. How does the current capacity increase?

   a) Capital sector shipments

   b) Goods sector shipments

   c) Depreciation

   d) None of the above

10. How does current capacity decrease?

   a) Capital sector shipments

   b) Goods sector shipments

   c) Depreciation

   d) None of the above

11. Depreciation consists of:

   a) The consumption of capital goods by the goods sector

   b) Lost orders in transit to the goods sector

   c) Capital orders produced for the capital sector's consumption

   d) Reduction in current capacity from wear and tear

12) Does depreciation reduce the capacity of the capital sector?

   a) Yes

   b) No

   c) Not applicable

13. The capital sector:

a) Produces goods to be consumed by the capital sector

b) Produces goods to be consumed by the goods sector

c) Consumes the depreciated material

d) All of the above

e) A and B only

14. Goods leave the production system via:

a) Shipment to the goods sector

b) Shipment to the capital sector

c) Depreciation

d) A and C only

15. If the capital sector is running at full capacity, and the goods sector demand increases, in the next period of play, meeting the increase in demand will cause:

a) The capital sector demand to increase

b) Underproduction

c) A backlog in the goods sector shipments

d) All of the above

e) None of the above

16. If the current capacity is larger than the total backlogs, what will occur to the game score?

a) It will go up

b) It will go down

c) It will remain the same

Continue on to next page

In this section, match an answer to the following questions. Note there are more answers than there are questions. You may refer to the attached diagram of the game board to assist you with some of the questions.

Questions:

1. What is represented by the field: Capital Stock Order Placement?

Answer: _____

2. What is represented by the field: Goods Sector Demand?

Answer: _____

3. How do you obtain desired production?

Answer: _____

4. How does the simulation calculate the current capacity?

Answer: _____

5. What is represented by the current capacity field?

Answer: _____

6. What factors are involved in calculating the allocation of orders?

Answer: _____

7. What factors are involved in computing your score?

Answer: _____

8. How is the current capacity proportioned?

Answer: _____

Answers:

A) Add goods sector capacity plus capital sector capacity

B) Ratio of capital sector backlog to goods sector backlog

C) Add capital sector backlog plus goods sector backlog (total backlogs)

D) Cumulative total of previous year's over/under production divided by the game period

E) Ratio of goods sector backlogs to depreciation

F) Goods order placements

G) Current capacity and total backlog

H) Capital sector demand

I) Total production capability

J) Depreciate capital stock, then add capital sector shipments to current capacity

K) Ratio of current capacity to depreciation

Answers to Knowledge Survey

(not provided to participants)

Multiple-Choice Questions		Matching Questions
1. B	9. A	1. H
2. E	10. C	2. F
3. C	11. D	3. C
4. B	12. A	4. J
5. A	13. B or E	5. I
6. E	14. D	6. G
7. D	15. D	7. D
8. B	16. A	8. B

APPENDIX H

Self-Assessment Survey Cover-Sheet

Decisions Within Complex Systems: An Experimental Approach Using the STATEGEM-2 Computer Game

Researcher: J. Robert Bois

This survey has been approved by the Institutional Review Board

State University of New York, Albany

Dear Participant,

You have just taken part in a voluntary experimental study of decision making within a dynamic system. Your personal results will always be kept totally confidential (known only by you and the researcher). The combined results of this study will assist other researchers, as well as decision makers in complex environments, to better understand the nature of the dynamic decision-making process. At this time, you are being asked to participate in a voluntary written survey. It is designed to assess the experiment and to gather feedback from you on your perceptions as a participant. The answers you provide here are expected to only enhance the research findings. The same rule of confidentiality expressed as being part of the experiment also applies to this survey.

Additionally, you are reminded that you are participating freely and that you are under no obligation to answer any or all of the questions. This survey should take, on average, no more than five minutes to complete. Thank you for your participation.

Please turn the page and be sure to read all instructions before beginning.

Self-Assessment Survey

Participant Number ___________

Please consider the following statements carefully. After each statement, circle the answer that best reflects your opinion. Would you say you strongly agree with the statement, agree, are neutral, disagree, or strongly disagree? Mark your answers accordingly on the scale for each question. As a reminder, you should answer each question as truthfully as possible. There are no wrong answers unless you are not being completely honest with yourself. Please go to the first question and begin.

Circle your response to each question.

1. Regarding *this survey*, I fully understand all that is required of me from the instructions.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. Regarding the *experiment*, I fully understood all that was required of me from the instructions.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I did my best in performing during the experiment.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. The experiment took too much time to complete.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. During the experiment, I sometimes forgot what I was supposed to do.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

6. Time constraints/pressures made me hurry during my responses.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

7. When provided with a set of decision cues to follow, I followed them all the time.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

8. I found that the *knowledge survey* was very difficult to accomplish.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

9. The required tasks were easy to understand.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

10. There were times when I found myself bored with completing the tasks.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

11. I am very interested in the outcomes of this research project.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

12. I gave this experiment "my all." I performed exactly in the manner prescribed in the verbal and written instructions.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Self-Assessment Survey Request for Comments

In the space provided below, please write any comments you would like to have known to the researcher. If you need more space, please use the reverse side of this page. Additionally, if in the future you would like to get in touch with the researcher; his contact information is provided at the bottom of this form.

You have just taken part in a voluntary written survey regarding your participation in an experimental study of decision making within a dynamic system. I would like to remind you that your answers and comments in this survey are to be kept totally confidential by the researcher. As an added note, you are also asked to not discuss this survey with any other participant until after all data collection is complete. If you have any further questions, please contact the researcher.

A sincere thank you for taking the time to participate in this study.

Researcher:
J. Robert Bois
boisj@nycap.rr.com
(518) 877-8781

Appendix I

Variables Collected and SPSS Column Codes

1. partcpnt : Participant Number
2. group : Group Number
3. gend : Gender : 1 = male; 2 = female
4. age : Age in years
5. grad : SUNY Status : 1 = undergraduate student; 2 = graduate student
6. yr : Year : 1 = junior; 2 = senior; 3 = graduate
7. prog : Program or Major : 1 = INFPhD; 2 = PADPhD; 3 = MPA; 4 = MPP; 5 = Marketing; 6 = Business Administration; 7 = Other
8. exp : Years professional experience
9. time : Time slot : 1 = 9am; 2 = 12pm; 3 = 3pm; 4 = 6pm
10. tut_t : Time on tutorial
11. prac_t : Time on practice
12. test_t : Time on test
13. game_t : Time on game
14. tt : Total time for experiment
15. t1 : Trial 1 score
16. Log10T1 : Base 10 Logarithmic conversion of T1 score
17. t2 : Trial 2 score
18. Log10T2 : Base 10 Logarithmic conversion of T2 score
19. ta : Trail average
20. Log10TA : Base 10 Logarithmic conversion of TA score
21. Delta : Change in score by subtracting Log10T2 from Log10T1
22. ts : Test score
23. q1 : Test question 1 (0 = incorrect response; 1 = correct response)
24. q2 : Test question 2
25. q3 : Test question 3
26. q4 : Test question 4
27. q5 : Test question 5
28. q6 : Test question 6
29. q8 : Test question 8
30. q9 : Test question 9
31. q10 : Test question 10
32. q11 : Test question 11

33. q12 : Test question 12
34. q13 : Test question 13
35. q14 : Test question 14
36. q15 : Test question 15
37. q16 : Test question 16
38. m1 : Matching question 1 (0 = incorrect response; 1 = correct response)
39. m2 : Matching question 2
40. m3 : Matching question 3
41. m4 : Matching question 4
42. m5 : Matching question 5
43. m6 : Matching question 6
44. m7 : Matching question 7
45. m8 : Matching question 8
46. sa1 : Self-assessment survey question 1
47. sa2 : Self-assessment survey question 2
48. sa3 : Self-assessment survey question 3
49. sa4 : Self-assessment survey question 4
50. sa5 : Self-assessment survey question 5
51. sa6 : Self-assessment survey question 6
52. sa7 : Self-assessment survey question 7
53. sa8 : Self-assessment survey question 8
54. sa9 : Self-assessment survey question 9
55. sa10 : Self-assessment survey question 10
56. sa11 : Self-assessment survey question 11
57. sa12 : Self-assessment survey question 12
58. sacom : Self-assessment survey comment : 0 = No, 1 = Yes

APPENDIX J

Improved Richardson and Rohrbaugh Rule Card

As the manager of the STRATEGEM-2 economy, you have taken it upon yourself to hire a very reputable economic consultant to assist you with your decisions. This person has determined that if you are to follow the formula in the box on the **reverse side** of this card, you will most likely receive an outstanding score for the game. You are reminded by this professional that although you are not required to heed the advice given, you must remain patient and diligent with using the formula (*use the reverse side for success!)*

Example on using the decision aide in year zero of the game:

1. Take the current depreciation of 50 units and multiply it times 2 (for 100).
2. Add to that the shortfall* (currently 0) and divide by 2 (which equals 0).
3. Then subtract the current capital backlog (*not total backlog*) of 50.
4. This produces an order of 50 capital units for Year 0.
   - When orders end in a "5," round UP to the nearest 10.
   - If orders compute to less than zero, use zero as your order.

* Shortfall = (total backlogs - current capacity) - always use the true value of this number, (either positive or negative). For example: *If this number computes to less than zero, then use that negative number and therefore, subtract it from Depreciation.*

1. Plan in advance to replace depreciation loss

   (DEPRECIATION X 2) = _______

2. Shortfall: Reconcile total backlogs with current capacity

   add or subtract (SHORTFALL* ÷ 2) = (+/ -) ______

3. Adjust for prior orders not yet filled

   subtract (CAPITAL BACKLOG *not Total Backlog*) = - _______

Total Orders = _______

(When Total Orders compute to a value of less than zero, use zero)

* Shortfall = (total backlogs - current capacity) - always use the true value of this number, (either positive or negative). For example: *If this number computes to less than zero, then use that negative number and therefore, subtract it from Depreciation.*